

THE SIGNED LOOP APPROACH TO THE ISING MODEL: FOUNDATIONS AND CRITICAL POINT

WOUTER KAGER, MARCIN LIS, AND RONALD MEESTER

ABSTRACT. The signed loop method is a beautiful way to rigorously study the two-dimensional Ising model with no external field. In this paper, we explore the foundations of the method, including all details that have so far been neglected or overlooked in the literature. We also demonstrate how the method can be applied to prove some classical results about the Ising model on the square lattice. This in turn leads to new explicit formulae for the free energy density and two-point functions in terms of sums over loops, valid all the way up to the self-dual point. It follows that the self-dual point is critical both for the behaviour of the free energy density, and for the decay of the two-point functions.

1. INTRODUCTION AND MAIN RESULTS

The Ising model [11] is probably the most studied model from statistical mechanics, and has been approached using many different ideas and from various angles. Most famous are the original algebraic methods of Onsager and Kaufman [14, 15, 18], used by them to compute the free energy and study correlations (see also [19]). In the fifties, Kac and Ward [12] proposed a very interesting alternative method for approaching the Ising model, based on studying configurations of signed loops. This method is combinatorial in nature, and as such often referred to as the *combinatorial method*. However, in order to avoid confusion with other approaches which also include combinatorial aspects (for instance, Kasteleyns dimer approach as exposed in [17]), we prefer to refer to the method as the *signed loop method*.

Over the years, a number of articles developing the Kac–Ward method have appeared in the physics literature, of which the most relevant are [3, 9, 21, 22, 25]. However, from a mathematical point of view, these papers, as well as the recent review [5], leave a lot to be desired in terms of mathematical rigour and technical details. Moreover, doubts have been cast on the very validity of the whole method to begin with, not only in the years following the Kac–Ward paper, but still relatively recently by Dolbilin et al. [6], who rightly pointed out an error in Vdovichenko’s article [25] (reproduced in [16, §151]) on the signed loop method.

Date: September 20, 2012.

2010 Mathematics Subject Classification. 82B20 (Primary) 60C05 (Secondary).

Key words and phrases. Ising model, signed loops, critical point, free energy density, two-point functions.

With the present paper, we aim to remove these doubts and deficiencies once and for all, by providing complete, rigorous and detailed proofs of the combinatorial identities central to the signed loop method, in great generality. In particular, we show that the error in Vdovichenko's article can be corrected by including an additional factor (the *multiplicity* defined below) into the loop weights.

The main objects of study in the signed loop method are graph generating functions. As we shall see, the one-point and two-point functions of the Ising model on \mathbb{Z}^2 , as well as the partition function, can all be expressed in terms of graph generating functions. As a result, the signed loop method can be used to prove both classical and new results for the Ising model. In particular, we show that the method leads to a proof of the fact that the self-dual point is in fact critical, both in terms of spontaneous magnetization, and as the only point where the free energy density is not analytic.

Needless to say, there are other approaches to the Ising model which appeal to mathematicians, and to probabilists in particular. We mention the approach of Aizenman, Barsky and Fernández [1], who prove sharpness of the phase transition using differential inequalities. We also mention the fermionic observables originally introduced by Smirnov [24] to study the Ising model at criticality, but which have been used in [2] in an interesting way for yet another derivation of the value of the critical point.

This paper is organised as follows. We present our key identities concerning graph generating functions and signed loops in Section 1.1, and we present our results for the Ising model on \mathbb{Z}^2 in Section 1.2. The proofs of our combinatorial identities are given in Section 2, and the proofs of our results for the Ising model can be found in Section 3.

1.1. The combinatorial identities. Although our applications are in \mathbb{Z}^2 , we will occasionally have to augment the graph on which we work. Therefore, it will be necessary to study the signed loop method on a more general graph structure, which is also of interest in itself for further applications. Hence, our starting point is a finite graph $G = (V, E)$, embedded in the plane, with vertex set V and edge set E . We identify G with its embedding. We assume G does not have multiple edges, but we do not assume that G is planar. For convenience, although this is not strictly necessary, we assume that the embedding is such that edges are straight line segments, and that except for the vertices at the two endpoints, no other vertices lie on an edge.

For practical reasons, we will think of edges as unordered pairs of vertices $\{u, v\}$, $u, v \in V$. For brevity, we will also write uv instead of $\{u, v\}$ for the edge between u and v (note that $uv = vu$). Since we do not assume that G is planar, it is possible that two edges intersect in a point which is not a vertex. In this case, we say that the two edges *cross* each other.

We call a subset F of E *even* if every vertex in the subgraph (V, F) of G has even degree (the empty set is also even). Our main object of study in this section will be a type of generating function for these even subgraphs.

To define this generating function, we assume that for every edge $uv \in E$, we are given an *edge weight* $x_{uv} \in \mathbb{R}$. We write $x = (x_{uv})_{uv \in E}$ for the vector of all edge weights. Furthermore, we denote by C_F the total number of unordered pairs of edges in F that cross each other. The generating function $Z(x)$ of even subgraphs of G is then defined as

$$(1.1) \quad Z(x) = \sum_{\text{even } F \subseteq E} (-1)^{C_F} \prod_{uv \in F} x_{uv}.$$

If the graph G is planar, then we can embed G in such a way that there are no edges that cross each other, so that $C_F = 0$ for all F . Note, however, that $Z(x)$ depends on the embedding, and that two different embeddings of the same (abstract) graph can lead to different functions $Z(x)$. Since we identify G with its embedding, this last fact does not concern us here.

The signed loop method is a way to compute the generating function $Z(x)$ by expressing it in terms of a sum over all possible *loops* in the graph G . Loosely speaking, a loop ℓ is a non-backtracking path passing through vertices $v_0, v_1, \dots, v_n = v_0$. It has a *sign* $\text{sgn}(\ell)$ which is -1 if its total winding angle is an even multiple of 2π , and $+1$ otherwise, and a *multiplicity* $m(\ell)$ counting how often the loop “repeats” itself (precise definitions will be given below). We associate with each loop $\ell = (v_0, \dots, v_{n-1})$ a signed *weight*

$$(1.2) \quad w(\ell; x) = \frac{\text{sgn}(\ell)}{m(\ell)} \prod_{i=0}^{n-1} x_{v_i v_{i+1}},$$

and we claim that with this choice of loop weights, the following holds:

Theorem 1.1. *For $uv \in E$, let d_{uv} denote the maximum of the degrees of u and v in the graph G . If $|x_{uv}| < (d_{uv} - 1)^{-1}$ for all $uv \in E$, then*

$$(1.3) \quad Z(x) = \exp\left(\sum_{\ell \text{ in } G} w(\ell; x)\right).$$

Under the condition of Theorem 1.1, the loop weights are absolutely summable, but we shall see that under a significantly weaker condition, we still obtain an equality similar to (1.3) if we sum the loop weights in order of increasing length of the loops. Before we explain this, we will first give the precise definitions of loops in G , and of their signs and multiplicities.

We start by introducing the notion of a *path*, since a loop is a specific type of path. A path of n steps in G is a sequence $(v_0, v_1, \dots, v_{n-1}) \in V^n$ such that $v_i v_{i+1} \in E$ for every i satisfying $0 \leq i \leq n-2$, and $v_{i+2} \neq v_i$ for every i such that $0 \leq i \leq n-3$ (paths are *non-backtracking*). If the path (v_0, \dots, v_{n-1}) satisfies $v_0 v_{n-1} \in E$, $v_0 \neq v_{n-2}$ and $v_1 \neq v_{n-1}$, then it is also called a *rooted loop* of n steps, rooted at v_0 . In this case, all rotations of the sequence (v_0, \dots, v_{n-1}) are also rooted loops of n steps, and so are all the rotations of the reverse sequence $(v_{n-1}, v_{n-2}, \dots, v_0)$.

We think of all these rooted loops as alternative representations of the same unrooted loop, and we want to choose one of these representations

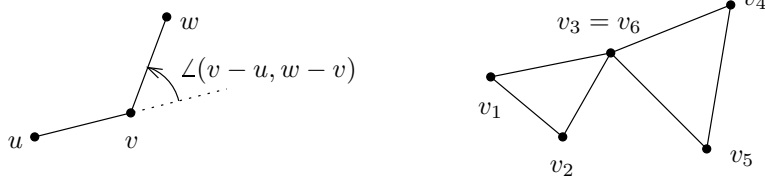


FIGURE 1. The turning angle from the vector $v - u$ to the vector $w - v$ (left). The loop $(v_1, v_2, v_3, v_4, v_5, v_6)$ on the right has sign -1 , the loop $(v_1, v_2, v_3, v_5, v_4, v_6)$ has sign $+1$.

as the one identifying the unrooted loop uniquely. To do so, we order the vertices of G lexicographically by their coordinates in the plane. Then there is a unique sequence (v_0, \dots, v_{n-1}) which is lexicographically smallest among the collection consisting of all its rotations, and all rotations of the reverse sequence (v_{n-1}, \dots, v_0) . Henceforth, when we speak of a *loop* (without the qualification *rooted*), we shall mean such a sequence which is smallest among the corresponding collection of permutations.

If $\ell = (v_0, \dots, v_{n-1})$ is a rooted or unrooted loop, we shall make the identification $v_j \equiv v_{j \bmod n}$ for all $j \in \mathbb{Z}$. We say that a loop ℓ is *edge-disjoint* if $v_i v_{i+1} \neq v_j v_{j+1}$ for all $i, j \in \{0, \dots, n-1\}$ with $i \neq j$. If ℓ is not edge-disjoint, it might be the case that the sequence (v_0, \dots, v_{n-1}) is periodic, in which case we call ℓ a *periodic loop*. The *multiplicity* of ℓ , denoted by $m(\ell)$, is its number of steps divided by its smallest period. In particular, the multiplicity of every nonperiodic loop is 1.

The *sign* of a loop is defined in terms of its *winding angle*, as follows. Given two distinct edges uv and vw , we define $\angle(v - u, w - v)$ as the turning angle in the plane from the vector $v - u$ to $w - v$, see Figure 1 (left). The winding angle $\alpha(\ell)$ of a loop $\ell = (v_0, \dots, v_{n-1})$ is simply the sum of all turning angles along the loop, that is,

$$(1.4) \quad \alpha(\ell) = \sum_{i=0}^{n-1} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}).$$

We now define the sign $\text{sgn}(\ell)$ of ℓ as

$$(1.5) \quad \text{sgn}(\ell) = -\exp\left(\frac{i}{2}\alpha(\ell)\right).$$

Observe that the winding angle of every loop is a multiple of 2π (here we use the fact that the edges of G are straight line segments), hence the sign of a loop is either $+1$ or -1 .

Remark. If a loop is edge-disjoint, it follows from Whitney's formula [26] that the sign of the loop is -1 if the loop crosses itself an odd number of times, and $+1$ if it crosses itself an even number of times (see Figure 1). For loops that are not edge-disjoint, it may not be so clear what is meant by a

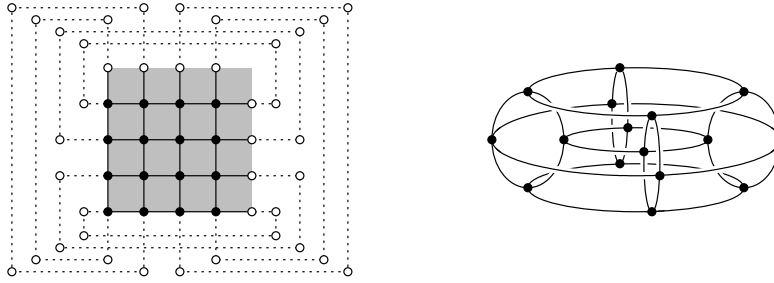


FIGURE 2. A square lattice wrapped around a torus (right) and a representation of it in the plane (left). The gray square corresponds to the torus, the dotted lines and open circles are the additional edges and vertices used for the embedding.

“crossing”, but definition (1.5) makes sense for both kinds of loop. Readers familiar with the recent work of Smirnov will notice the similarity between definition (1.5) and the definition of the fermionic observable in [24], which is also in terms of winding angles.

Note that there are infinitely many loops in G (unless G has no cycles), since loops can have any number of steps. In what is to follow, it becomes essential to sum over loops in the correct order; we will collect loops of a given length together, and sum in order of increasing length. However, we will not always identify the *length* of a loop with its *number of steps*.

For instance, in our applications to the Ising model on \mathbb{Z}^2 , we will have reasons to consider loops on a square lattice wrapped on a torus. However, for the signed loop method and our Theorems 1.1 and 1.2, it is essential to work on a graph G embedded in the plane. Instead of working on the torus directly, we must therefore work on a suitable representation of it in the plane. To keep the computations simple, the most convenient representation will be a rectangular piece of \mathbb{Z}^2 , to which some extra edges and vertices are added to connect opposite sides, as illustrated in Figure 2.

Some of the edges in the constructed graph G now do not correspond to edges that can be traversed by a loop on the torus, and therefore, steps taken along these added edges should not contribute to the length of a loop. We call these extra edges *additional* edges, and denote the set of additional edges by E_A . The edges in $E \setminus E_A$ are called *representative* edges, since they do correspond to edges in the represented graph. In general, these considerations lead us to allow the edge set E of the graph G we work on to be divided into a set E_A of additional edges and a set $E \setminus E_A$ of representative edges. For reasons that will become clear, we must assume that the set E_A is such that the graph (V, E_A) is free of cycles.

Additional edges will also play another role. In our study of correlations in the Ising model, they are used to artificially connect two spin variables

whose correlation we want to study. Again, also in this role, we do not want the additional edges to contribute to the length of a loop.

For these reasons, we will measure the length of a loop in terms of the number of *representative* edges it traverses, ignoring any steps taken along *additional* edges. To be precise, we formally define the *length* $r(\ell)$ of the loop $\ell = (v_0, \dots, v_{n-1})$ as the number of i in $\{0, \dots, n-1\}$ such that $v_i v_{i+1} \in E \setminus E_A$. Note the distinction between the *length* of a loop and its *number of steps*. By \mathcal{L}_r we denote the collection of all loops of length r , and we write $\mathcal{L} = \bigcup_{r=1}^{\infty} \mathcal{L}_r$ for the collection of all loops. We can now write

$$(1.6) \quad f_r(x) = \sum_{\ell \in \mathcal{L}_r} w(\ell; x)$$

for the sum of weights of all loops of length r . Under the condition of Theorem 1.1, the loop weights are absolutely summable. This implies in particular that $Z(x)$ equals $\exp(\sum_r f_r(x))$, but we shall see that the latter policy of summing the loop weights by length can already be used under a significantly weaker condition.

To formulate this condition, we need to introduce the *transition matrix* $\Lambda(x)$. If uv is an edge of G , then by \vec{uv} we will denote the *directed* edge from u to v . The matrix $\Lambda(x)$ will be indexed by the *directed representative* edges of G . Given two directed representative edges \vec{uv} and \vec{wz} , we say that v is *linked* to w if either $v = w$, or there exists a sequence of distinct *additional* edges $v_1 v_2, v_2 v_3, \dots, v_{n-1} v_n$ such that $v = v_1$ and $v_n = w$. In the former case, if $v = w$ and $u \neq z$, we write

$$\angle(\vec{uv}, \vec{wz}) = \angle(v - u, z - w)$$

for the turning angle from \vec{uv} to \vec{wz} . In the latter case, the sequence (v_1, \dots, v_n) is a path (the *chain*) linking v to w , passing through additional edges only, and we say that “ $v \rightsquigarrow w$ via (v_1, \dots, v_n) ”. By our assumption that the additional edges form no cycles, there can be at most one such path. Hence, without ambiguity, if v is linked to w in this way, we can define

$$\begin{aligned} \angle(\vec{uv}, \vec{wz}) &= \angle(v - u, v_2 - v) \\ &\quad + \sum_{i=1}^{n-2} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}) + \angle(w - v_{n-1}, z - w). \end{aligned}$$

The transition matrix $\Lambda(x)$ is defined as follows. Write $\Lambda_{\vec{uv}, \vec{wz}}(x)$ for the entry of the matrix with row index \vec{uv} and column index \vec{wz} . Then

$$(1.7) \quad \Lambda_{\vec{uv}, \vec{wz}}(x) = \begin{cases} x_{uv} e^{i\angle(\vec{uv}, \vec{wz})/2} & \text{if } v = w \text{ and } u \neq z; \\ x_{uv} \prod_{i=1}^{n-1} x_{v_i v_{i+1}} e^{i\angle(\vec{uv}, \vec{wz})/2} & \text{if } v \rightsquigarrow w \text{ via } (v_1, \dots, v_n); \\ 0 & \text{otherwise.} \end{cases}$$

Let $\lambda_i(x)$, $i = 1, 2, \dots, 2|E \setminus E_A|$, denote the eigenvalues of $\Lambda(x)$. We will show that if $\max_i |\lambda_i(x)| < 1$, then the $f_r(x)$ are absolutely summable. This leads to an identity which forms the core of the signed loop method:

Theorem 1.2. *If $\max_i |\lambda_i(x)| < 1$, then*

$$(1.8) \quad Z(x) = \exp\left(\sum_{r=1}^{\infty} f_r(x)\right) = \sqrt{\det(\mathbf{I} - \Lambda(x))}.$$

An alternative proof of the (Kac–Ward) formula $Z(x)^2 = \det(\mathbf{I} - \Lambda(x))$ for the case of a planar graph G with straight edges, and without consideration of the possibility of including additional edges, was given in [7]. Our results are not only more general, but also express $Z(x)$ in terms of loops and loop configurations. Both these facts are used in an essential way in our applications to the Ising model. Extensions of the Kac–Ward formula to graphs embedded in surfaces of higher genus are studied in [4].

1.2. Applications to the Ising model. In this section, we consider applications of Theorem 1.2 to the two-dimensional Ising model. We focus our attention here on studying the phase transition of the isotropic Ising model on the square lattice \mathbb{Z}^2 , in terms of the behaviour of the free energy density and of the two-point functions. Similar results as the ones we obtain here for the square lattice can also be obtained for the triangular and hexagonal lattices with the signed loop method. In fact, the method applies to the Ising model on even more general graphs, and allows one to study general k -point functions. We will go into these issues in a subsequent paper.

In our applications to the Ising model, $\|\cdot\|$ will denote the L_1 -norm in \mathbb{Z}^2 or \mathbb{R}^2 . Thus, $\|u - v\|$ measures the graph distance between vertices u and v of \mathbb{Z}^2 or \mathbb{Z}^{2*} , the dual of \mathbb{Z}^2 . The notation $\|\cdot\|_{\infty}$ is used for the maximum norm on the relevant space, and $o = (0, 0)$ denotes the origin in \mathbb{Z}^2 .

We shall start with a brief review of the definition of the Ising model and of the low- and high-temperature expansions, through which the connection with the graph generating function $Z(x)$ of Section 1.1 will be made. We consider the model on finite rectangles in the square lattice \mathbb{Z}^2 . We call a graph $G = (V, E)$ a *rectangle in \mathbb{Z}^2* if the vertex set V consists of all points of \mathbb{Z}^2 contained in a rectangle $[a, b] \times [c, d]$ in \mathbb{R}^2 , and for $u, v \in V$, the edge set E contains the edge uv if and only if $\|u - v\| = 1$. We say that vertices u and v are *neighbours* if $uv \in E$, and we define the *boundary* ∂G of G as the set of vertices that have less than 4 neighbours in G . By $G^* = (V^*, E^*)$ we shall denote the *weak dual* graph of G , i.e. the rectangle in \mathbb{Z}^{2*} whose vertices are the centres of the faces of G (see Figure 3, left).

To define the Ising model on G , we associate with each vertex $v \in V$ a *spin variable* σ_v , taking values in $\{-1, +1\}$. Thus, the space of all possible spin configurations is $\Omega = \{-1, +1\}^V$. Sometimes we shall impose *positive boundary conditions*, by which we mean that we restrict ourselves to the set

of spin configurations

$$\Omega^+ = \{\sigma \in \Omega: \sigma_u = +1 \text{ if } u \in \partial G\}.$$

In contrast, when we speak of *free boundary conditions*, we work with the unrestricted set of spin configurations $\Omega^{\text{free}} = \Omega$.

To each spin configuration $\sigma = (\sigma_v)_{v \in V}$ we associate an *energy* given by the Hamiltonian

$$H(\sigma) = - \sum_{uv \in E} \sigma_u \sigma_v.$$

If we interpret the spin variables as magnetic moments, this Hamiltonian describes ferromagnetic interactions, since the energy is lower the more pairs of neighbouring spins are in the same state. According to the formalism of statistical mechanics, the equilibrium probability distribution of the spin configurations at the inverse temperature β is now given by the Gibbs–Boltzmann distribution

$$(1.9) \quad P_\beta^\square(\sigma) = \frac{1}{Z_\beta^\square} e^{-\beta H(\sigma)} = \frac{1}{Z_\beta^\square} \prod_{uv \in E} e^{\beta \sigma_u \sigma_v}, \quad \sigma \in \Omega^\square,$$

where $\square \in \{\text{free}, +\}$ stands for the imposed boundary condition, and Z_β^\square , the *partition function*, is given by

$$(1.10) \quad Z_\beta^\square = \sum_{\sigma \in \Omega^\square} e^{-\beta H(\sigma)} = \sum_{\sigma \in \Omega^\square} \prod_{uv \in E} e^{\beta \sigma_u \sigma_v}.$$

1.2.1. Low- and high-temperature expansions. The partition function of the Ising model is closely related to the graph generating function $Z(x)$ from Section 1.1. To see this, we now consider the so-called *low-temperature* and *high-temperature* expansions. More details on these expansions and the related duality of the Ising model can be found in [23, Section II.7]. For our purposes, the low-temperature expansion is best considered in the case of positive boundary conditions. It is not difficult to see that in this case, there is a 1–1 correspondence between the even subgraphs of the weak dual G^* and the spin configurations in Ω^+ : given $\sigma \in \Omega^+$, one obtains the corresponding even subset $F(\sigma)$ of E^* by including the edge dual to uv in $F(\sigma)$ if and only if $\sigma_u \neq \sigma_v$, for every $uv \in E$. See Figure 3 (left) for an illustration.

Note that by this correspondence, if $F \subset E^*$ is even, then every edge in F separates two spins that have opposite sign. Since the energy increases by 2 if $\sigma_u \sigma_v$ changes from +1 to −1, this means that every edge $uv \in F$ comes with a “cost” of $\exp(-2\beta)$. It follows that we can write

$$(1.11) \quad P_\beta^+(\sigma) = \frac{\exp(\beta|E|)}{Z_\beta^+} \prod_{uv \in F(\sigma)} e^{-2\beta}, \quad \sigma \in \Omega^+,$$

where

$$(1.12) \quad Z_\beta^+ = \exp(\beta|E|) \sum_{\text{even } F \subset E^*} \prod_{uv \in F} e^{-2\beta}.$$

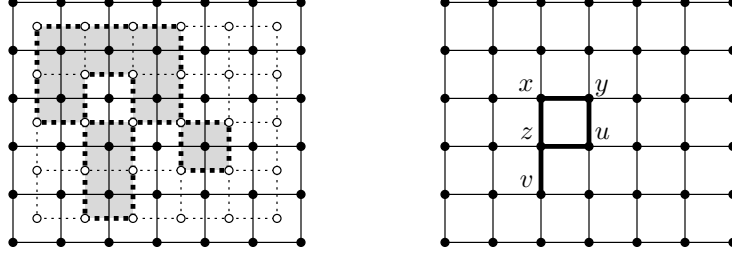


FIGURE 3. Left: the graph G and its weak dual G^* , with an even subgraph of G^* marked by bold dashed edges. The spins in the gray squares have value -1 , the rest have value $+1$. Right: the subgraph of G drawn with bold edges contributes $\sigma_x^2 \sigma_y^2 \sigma_z^3 \sigma_u^2 \sigma_v^2 (\tanh \beta)^5$ in the high-temperature expansion.

This is the low-temperature expansion of the partition function for positive boundary conditions. Observe that up to the factor $\exp(\beta|E|)$, this expansion takes exactly the form (1.1) of the graph generating function $Z(x)$ for the dual graph G^* (in which no edges cross each other), if we set the edge weights of all dual edges $uv \in E^*$ equal to $x_{uv} = \exp(-2\beta)$.

We now turn to the high-temperature expansion, for which we impose free boundary conditions. The expansion will be over even subgraphs of the graph G , rather than of the dual G^* . Unlike in the low-temperature expansion, these subgraphs do not have a clear geometric interpretation. We will therefore take some time to explain how they arise.

The high-temperature expansion starts from equation (1.10) and the observation that $\sigma_u \sigma_v$ can only take the values -1 or $+1$. Since $\exp(\pm\beta) = \cosh \beta \pm \sinh \beta$, we see that

$$Z_\beta^{\text{free}} = (\cosh \beta)^{|E|} \sum_{\sigma \in \Omega^{\text{free}}} \prod_{uv \in E} (1 + \sigma_u \sigma_v \tanh \beta).$$

The next step is to expand the product over $uv \in E$. Each term in the expansion will be a product of factors obtained by choosing for each edge uv whether 1 is taken as a factor, or $\sigma_u \sigma_v \tanh \beta$. The expansion can therefore be written as the sum of the products of factors over all choices one can make for the factors coming from each edge uv . We can represent each choice graphically by removing the edge uv if we choose the factor 1 for this edge, and keeping uv if we choose the factor $\sigma_u \sigma_v \tanh \beta$. This gives a 1–1 correspondence between all terms in the expansion, and all $F \subset E$ (not just the even ones). See also Figure 3 (right).

Using this correspondence, and then interchanging the order of summation over σ and F , we may now write the partition function as

$$Z_\beta^{\text{free}} = (\cosh \beta)^{|E|} \sum_{F \subset E} \sum_{\sigma \in \Omega^{\text{free}}} \prod_{u \in V} \sigma_u^{\deg(u, F)} \prod_{uv \in F} \tanh \beta,$$

where $\deg(u, F)$ denotes the degree of u in the graph (V, F) . Note that the sum over σ vanishes unless $\deg(u, F)$ is even for all $u \in V$, in which case the sum yields simply $2^{|V|}$. Therefore,

$$(1.13) \quad Z_\beta^{\text{free}} = 2^{|V|} (\cosh \beta)^{|E|} \sum_{\text{even } F \subset E} \prod_{uv \in F} \tanh \beta.$$

Again, up to a multiplicative constant, the expansion takes exactly the form (1.1) of the graph generating function $Z(x)$, this time for the graph G , if we set the edge weights equal to $x_{uv} = \tanh \beta$.

1.2.2. Results for the Ising model. The connection between the low- and high-temperature expansions and the generating function $Z(x)$ can be exploited to derive a number of results for the Ising model. In this paper, we focus on specific aspects of the phase transition in the isotropic Ising model on the square lattice. We shall study in particular the free energy density and the behaviour of the two-point functions.

In order to apply Theorem 1.2, we must obtain a bound on the spectral radius $\max_i |\lambda_i(x)|$ of $\Lambda(x)$. Our following result therefore provides the key to applying the signed loop method to the Ising model on \mathbb{Z}^2 :

Theorem 1.3. *For a finite rectangle G in \mathbb{Z}^2 with no additional edges,*

$$\max_i |\lambda_i(x)| \leq (\sqrt{2} + 1) \|x\|_\infty,$$

$$\text{and } |f_r(x)| \leq 2|V|(\sqrt{2} + 1)^r \|x\|_\infty^r.$$

Recall that we will take our edge weights to be respectively $\exp(-2\beta)$ and $\tanh \beta$ in the low- and high-temperature expansions. Thus, Theorem 1.3 says that we can apply the signed loop method for all temperatures up to the self-dual point, which is given by the value of β_c satisfying

$$(1.14) \quad \exp(-2\beta_c) = \tanh \beta_c = \sqrt{2} - 1.$$

We write β_c here for the self-dual point, because our results actually imply that β_c is the *critical* point of the Ising model in two senses: it is the only point where the free energy density is not analytic, and also the point where the behaviour of the two-point functions changes from staying bounded away from zero, to decaying exponentially with distance. A computation of the latter interpretation of the critical point was recently also performed in [2], using a very different method based on Smirnov's fermionic observable [24].

The reader may be surprised to learn that the signed loop method applies all the way up to the critical point, while classical arguments like the Peierls and Fisher arguments [8, 10, 20] stay far from it. What happens here is the following. If we simply count the number of non-backtracking (unsigned) loops in \mathbb{Z}^2 of n steps starting from a vertex v , this number grows like 3^n . However, if we take the signs of these loops into account, first counting all loops that have positive sign, and then subtracting those that have a negative sign, the obtained number only grows (in absolute value) like $(\sqrt{2} + 1)^n$. This reduction in growth rate is precisely enough to take us to the critical point.

We now turn to our results for the Ising model. First we consider the *Helmholtz free energy* F_β^\square , which is related to the partition function through

$$F_\beta^\square = -\frac{1}{\beta} \ln Z_\beta^\square.$$

The free energy is an important quantity, because it encodes all kinds of thermodynamic information about the system: derivatives of the free energy with respect to temperature are related to physical properties such as internal energy and specific heat. In particular, the formalism of statistical mechanics predicts that phase transitions coincide with discontinuities or other singularities in derivatives of the free energy. However, since the free energy of any finite system is always an analytic function of the parameters, such singularities can only occur in the infinite-volume limit.

We will therefore consider limits as our rectangles $G = (V, E)$ tend to \mathbb{Z}^2 (i.e. such that every finite subgraph of \mathbb{Z}^2 is eventually contained in G). Since now we no longer work on a fixed graph, we shall make the dependence on the graph explicit in our notation by writing $Z_{G,\beta}^\square$ for the partition function of the Ising model on G , and $F_{G,\beta}^\square$ for the free energy. It is a well-known fact that the infinite-volume limit of the free energy density, i.e.

$$\lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} F_{G,\beta}^\square = \lim_{G \rightarrow \mathbb{Z}^2} -\frac{1}{\beta|V|} \ln Z_{G,\beta}^\square =: f(\beta),$$

exists (which we will prove). It is also not difficult to show that the limit is the same for all boundary conditions, see for example [23, Section II.3].

We will use Theorem 1.2 to derive an explicit power series expression for $f(\beta)$. This implies directly that $f(\beta)$ is an analytic function of β on $(0, \beta_c) \cup (\beta_c, \infty)$, and moreover, it allows us to derive the formula for $f(\beta)$ which was first obtained by Onsager in [18]:

Theorem 1.4. *The free energy density $f(\beta)$ is an analytic function of β on $(0, \beta_c) \cup (\beta_c, \infty)$, where it is given by the Onsager formula*

$$-\frac{1}{\beta} \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln[4 \cosh^2 2\beta - 4 \sinh 2\beta (\cos \omega_1 + \cos \omega_2)] d\omega_1 d\omega_2.$$

The functions f and $u = \partial(\beta f)/\partial\beta$, which is the *internal energy density* of the system, are both continuous functions of β . However, Onsager has shown in [18] that the derivative of u with respect to temperature, which is the *specific heat*, diverges as β approaches β_c . This shows that β_c is critical for the behaviour of the free energy. But we would really like to show that β_c is also critical for the magnetic behaviour of the Ising model. Next, we therefore consider the one-point and two-point functions above and below β_c .

To simplify the notation, and following the physics literature, we write

$$(1.15) \quad \langle \sigma_u \rangle_{G,\beta}^\square = E_{G,\beta}^\square(\sigma_u), \quad \langle \sigma_u \sigma_v \rangle_{G,\beta}^\square = E_{G,\beta}^\square(\sigma_u \sigma_v),$$

where $E_{G,\beta}^\square$ denotes expectation with respect to the Ising measure $P_{G,\beta}^\square$. Our next result says that for positive boundary conditions, at all temperatures

below the critical point, the system has a net magnetization per unit volume in the infinite-volume limit:

Theorem 1.5. *For all $\beta \in (\beta_c, \infty)$ and fixed $u, v \in \mathbb{Z}^2$, the limits*

$$\lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \rangle_{G, \beta}^+ =: \langle \sigma_u \rangle_{\mathbb{Z}^2, \beta}^+ \quad \text{and} \quad \lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle_{G, \beta}^+ =: \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^+$$

both exist and are strictly positive. Furthermore, we have that

$$\lim_{\|u-v\| \rightarrow \infty} \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^+ = [\langle \sigma_o \rangle_{\mathbb{Z}^2, \beta}^+]^2 > 0.$$

For free boundary conditions, by symmetry under changing the values of all spins, $\langle \sigma_u \rangle_{G, \beta}^{\text{free}} = 0$, so the infinite-volume limit $\langle \sigma_u \rangle_{\mathbb{Z}^2, \beta}^{\text{free}}$ trivially exists and is 0. The next theorem addresses the behaviour of the two-point functions at high temperatures:

Theorem 1.6. *For all $\beta \in (0, \beta_c)$ and fixed $u, v \in \mathbb{Z}^2$, the limit*

$$\lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} =: \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^{\text{free}}$$

exists. Moreover, we have that

$$\lim_{\|u-v\| \rightarrow \infty} \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^{\text{free}} = [\langle \sigma_o \rangle_{\mathbb{Z}^2, \beta}^{\text{free}}]^2 = 0.$$

Our proofs show that in all cases in Theorems 1.5 and 1.6, the convergence is exponentially fast. Note the contrasting behaviour of the two-point functions above and below β_c : at low temperatures they stay bounded away from 0, while for high temperatures they decay to 0. This is an indication of the phase transition of the Ising model in terms of its magnetic behaviour.

However, we have used different boundary conditions above and below the critical point. Ideally, we would like to show that there is no magnetization in the high-temperature regime even for positive boundary conditions, and that at low temperatures, the two-point functions stay bounded away from zero for free boundary conditions as well. We believe that this can be shown using the signed loop method, but leave these issues for a subsequent paper.

2. PROOFS OF THE COMBINATORIAL IDENTITIES

We now turn to the proof of our main Theorems 1.1 and 1.2. Recall that here $G = (V, E)$ is a general finite graph embedded in the plane, potentially containing crossing edges or additional edges. The proof proceeds in a number of steps. In the first step, detailed in Section 2.1, we will identify each even subgraph (V, F) of G with a number of edge-disjoint collections of loops, and show that the sum of their weights yields precisely the contribution of F to the generating function $Z(x)$. In the second step, in Section 2.2, we will explain the conditions under which we can express $Z(x)$ in terms of $\sum_r f_r(x)$ and $\det(\mathbf{I} - \Lambda(x))$, under the assumption that the weights of all remaining configurations of loops in G of total length r cancel each other. The proof of this assumption, the last step, will be carried out in Section 2.3.

2.1. Expansion into collections of edge-disjoint loops. We will be concerned with crossings of loops and paths, and we need to carefully establish the relevant definitions first. More specifically, we will consider collections $\{\ell_1, \dots, \ell_s\}$ of loops with the properties that all loops ℓ_1, \dots, ℓ_s are edge-disjoint, and no two loops in the collection visit a common edge. We call these *edge-disjoint collections* of loops. Intuitively it may be clear what we mean by a crossing of such edge-disjoint loops, but some care is needed, so we will now give the precise definitions.

First, consider two paths (u, v, w) and (x, y, z) in G , and let A be the union of the two half-lines $\{v + t(u - v) : t \geq 0\}$ and $\{v + t(w - v) : t \geq 0\}$. We say that (u, v, w) and (x, y, z) *cross each other at the vertex v* if $v = y$ and the vertices x and z do not lie in the same infinite component of the complement of A in the plane.

Now let $\ell_1 = (u_0, \dots, u_{n-1})$ and $\ell_2 = (v_0, \dots, v_{m-1})$ be two loops that form an edge-disjoint pair $\{\ell_1, \ell_2\}$. By $C_V(\ell_1, \ell_2)$ we denote the number of pairs (i, j) , where $0 \leq i < n$ and $0 \leq j < m$, such that the paths (u_{i-1}, u_i, u_{i+1}) and (v_{j-1}, v_j, v_{j+1}) cross each other at u_i . We call $C_V(\ell_1, \ell_2)$ the number of *vertex crossings* between ℓ_1 and ℓ_2 . Similarly, we define the number of *edge crossings* between ℓ_1 and ℓ_2 , denoted $C_E(\ell_1, \ell_2)$, as the number of pairs (i, j) such that $0 \leq i < n$ and $0 \leq j < m$, and the edges $u_i u_{i+1}$ and $v_j v_{j+1}$ cross each other in G .

We also need to formally define the number of times a loop crosses itself, so consider an edge-disjoint loop $\ell = (v_0, \dots, v_{n-1})$. We define the number of *vertex self-crossings* of ℓ , denoted $C_V(\ell)$, as the number of pairs (i, j) , where $0 \leq i < j < n$, such that (v_{i-1}, v_i, v_{i+1}) and (v_{j-1}, v_j, v_{j+1}) cross each other at the vertex v_i . The number of *edge self-crossings* of ℓ , denoted $C_E(\ell)$, is defined as the number of pairs (i, j) such that $0 \leq i < j < n$, and the edges $v_i v_{i+1}$ and $v_j v_{j+1}$ cross each other in the graph G .

As we have already mentioned in the introduction, Whitney's formula [26] says that the sign of an edge-disjoint loop is -1 if the loop crosses itself an odd number of times, and $+1$ otherwise. In other words, we have

$$\text{sgn}(\ell) = (-1)^{C_V(\ell) + C_E(\ell)}$$

if ℓ is edge-disjoint. We now simply define the sign of an edge-disjoint collection of loops $\{\ell_1, \dots, \ell_s\}$ as

$$(2.1) \quad \text{sgn}\{\ell_1, \dots, \ell_s\} = \prod_{i=1}^s \text{sgn}(\ell_i) = (-1)^{\sum_{i=1}^s \{C_V(\ell_i) + C_E(\ell_i)\}}.$$

If $F \subset E$ is even, we can decompose F into an edge-disjoint collection of loops in such a way, that the union of all edges traversed by the loops is F (one way to find such a decomposition is given in the proof of Proposition 2.1 below). This decomposition is in general not unique. We write $\mathcal{D}(F)$ for the set of all possible edge-disjoint decompositions of F , and recall that C_F denotes the number of unordered pairs of edges in F that cross each other.

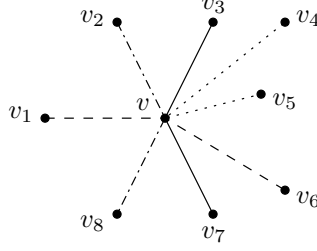


FIGURE 4. The neighbours v_1, v_2, \dots of a vertex v in an even subgraph (V, F) are ordered in a clockwise fashion around v . Two neighbours that are connected by edges drawn in the same line style are paired to each other (see the text).

Proposition 2.1. *For all even subsets F of E we have that*

$$\sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)} \text{sgn}\{\ell_1, \dots, \ell_s\} = (-1)^{C_F}.$$

Proof. Let $\{\ell_1, \dots, \ell_s\}$ be an edge-disjoint collection of loops. Since ℓ_1, \dots, ℓ_s are all closed loops in the plane, any two distinct loops ℓ_i and ℓ_j from the collection necessarily cross each other an even number of times. That is, $C_V(\ell_i, \ell_j) + C_E(\ell_i, \ell_j)$ is even for all $i \neq j$. Therefore,

$$\text{sgn}\{\ell_1, \dots, \ell_s\} = (-1)^{\sum_{1 \leq i \leq s} \{C_V(\ell_i) + C_E(\ell_i)\} + \sum_{1 \leq i < j \leq s} \{C_V(\ell_i, \ell_j) + C_E(\ell_i, \ell_j)\}}.$$

Furthermore, if $\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)$, then clearly the total number of edge crossings occurring among the loops must coincide with C_F , that is,

$$C_F = \sum_{1 \leq i \leq s} C_E(\ell_i) + \sum_{1 \leq i < j \leq s} C_E(\ell_i, \ell_j).$$

Hence, it suffices to prove that

$$(2.2) \quad \sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)} (-1)^{\sum_{1 \leq i \leq s} C_V(\ell_i) + \sum_{1 \leq i < j \leq s} C_V(\ell_i, \ell_j)} = 1.$$

Let V_F denote the set of vertices in V whose degree in the subgraph (V, F) is nonzero, and let $\deg(v, F)$ denote the degree of v in (V, F) . For $v \in V_F$ of degree $2k$, write v_1, \dots, v_{2k} for the endpoints of the edges in F that are incident to v . Assume that these vertices are ordered in a clockwise manner around v , starting from the lexicographically smallest one (see Figure 4). Denote by $\mathcal{P}_v(F)$ the collection of partitions of the vertices v_1, \dots, v_{2k} into sets of size 2. We call these partitions *pairings at the vertex v* . We write

$$\mathcal{P}(F) = \prod_{v \in V_F} \mathcal{P}_v(F),$$

and call an element of $\mathcal{P}(F)$ a *pairing* associated with the subgraph (V, F) .

We have a natural 1–1 correspondence between $\mathcal{P}(F)$ and $\mathcal{D}(F)$. Indeed, starting from any vertex $v \in V_F$ and any $i \in \{1, \dots, \deg(v, F)\}$, the pairing $\pi \in \mathcal{P}(F)$ defines a unique rooted loop (u_0, \dots, u_{n-1}) with the properties that $u_0 = v_i$, $u_1 = v$, and for all j , u_{j-1} is paired with u_{j+1} at the vertex u_j . Continuing this way, and replacing each rooted loop obtained by its unrooted counterpart, yields an edge-disjoint collection $\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)$. It is easy to see that this defines a bijective relation between $\mathcal{P}(F)$ and $\mathcal{D}(F)$.

Using this bijection, we can express the sum in (2.2) equally well as a sum over all pairings. More precisely, for $\pi \in \mathcal{P}(F)$ let π_v denote the pairing it induces at the vertex $v \in V_F$, and write $C_v(\pi_v)$ for the number of crossings at the vertex v introduced by this pairing. We call π_v *even* (*odd*) if $C_v(\pi_v)$ is even (odd). Note that (2.2) is equivalent to

$$\sum_{\pi \in \mathcal{P}(F)} (-1)^{\sum_{v \in V_F} C_v(\pi_v)} = \prod_{v \in V_F} \sum_{\pi_v \in \mathcal{P}_v(F)} (-1)^{C_v(\pi_v)} = 1,$$

from which we see that it suffices to prove that for all $v \in V_F$, the number of even pairings π_v exceeds the number of odd pairings π_v by 1.

We prove this by induction on the degree $2k$ of v . Write N_k^+ and N_k^- for the numbers of even, resp. odd, pairings for v of degree $2k$. For $k = 1$ we clearly have $N_k^+ = 1$ and $N_k^- = 0$. Now let $k > 1$, and suppose that we pair the vertex v_1 with v_i at v . Next pair the remaining $2k - 2$ neighbours v_j of v in all possible ways. For even i , there is an even number of j in between 1 and i , and therefore the pairing we obtain will be even if and only if the pairing of the remaining $2k - 2$ vertices is even (see Figure 4). Likewise, for odd i , the obtained pairing will be even if and only if the pairing of the remaining vertices is odd. Since we have k even values for i , and $k - 1$ odd values, this gives $N_k^+ = kN_{k-1}^+ + (k-1)N_{k-1}^-$ and $N_k^- = kN_{k-1}^- + (k-1)N_{k-1}^+$. Hence by the induction hypothesis, $N_k^+ - N_k^- = 1$. \square

From Proposition 2.1 we will now obtain our first main result. Recall that

$$Z(x) = \sum_{\text{even } F \subseteq E} (-1)^{C_F} \prod_{uv \in F} x_{uv}.$$

The case $F = \emptyset$ is treated separately: by convention it contributes 1 to the sum. Hence, Proposition 2.1 implies that

$$Z(x) = 1 + \sum_{\substack{\text{even } F \subseteq E: \\ F \neq \emptyset}} \sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)} \text{sgn}\{\ell_1, \dots, \ell_s\} \prod_{uv \in F} x_{uv}.$$

Recall that the multiplicity $m(\ell)$ of an edge-disjoint loop ℓ is 1. Therefore, using (2.1) and the definition (1.2) of the weight of a loop, we can write

$$Z(x) = 1 + \sum_{\substack{\text{even } F \subseteq E: \\ F \neq \emptyset}} \sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)} \prod_{i=1}^s w(\ell_i; x).$$

Since this is a finite sum, we do not need to worry about the order of summation, so we have

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{\substack{\text{even } F \subset E: \\ |F \setminus E_A| = r}} \sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)} \prod_{i=1}^s w(\ell_i; x).$$

If we now denote by \mathcal{D}_r the set consisting of all those edge-disjoint collections of loops $\{\ell_1, \dots, \ell_s\}$ for which the total length $\sum_{i=1}^s r(\ell_i)$ is r , we see that we have established the following theorem:

Theorem 2.2.

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{\{\ell_1, \dots, \ell_s\} \in \mathcal{D}_r} \prod_{i=1}^s w(\ell_i; x).$$

2.2. Extension to all loop configurations. In Theorem 2.2, we have expressed the generating function $Z(x)$ as a sum over all edge-disjoint collections of loops in G . In this section, we will see that if the edge weights are sufficiently small, we can drop the condition that the loops have to be edge-disjoint, and sum instead over all possible loop configurations in G . Here, a *loop configuration* is simply an ordered sequence (ℓ_1, \dots, ℓ_s) of loops; there is no condition that loops have to be edge-disjoint, nor that two loops in the configuration have to be distinct (i.e. it is allowed that $\ell_i = \ell_j$ for some $i \neq j$, which is why we work with ordered sequences of loops now).

Write \mathcal{C}_r for the collection of all loop configurations (ℓ_1, \dots, ℓ_s) satisfying $r(\ell_1) + \dots + r(\ell_s) = r$. Some of these loop configurations will consist of distinct loops that together form an edge-disjoint collection of loops. Let \mathcal{C}_r^* denote the subset of \mathcal{C}_r containing only these edge-disjoint loop configurations. Observe that if $\{\ell_1, \dots, \ell_s\}$ is an edge-disjoint collection of loops, then the corresponding loop configuration (ℓ_1, \dots, ℓ_s) has $s!$ permutations. Therefore, by Theorem 2.2 we already have that

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \sum_{(\ell_1, \dots, \ell_s) \in \mathcal{C}_r^*} \frac{1}{s!} \prod_{i=1}^s w(\ell_i; x),$$

but we claim that here we may sum over \mathcal{C}_r instead of \mathcal{C}_r^* :

Theorem 2.3.

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \sum_{(\ell_1, \dots, \ell_s) \in \mathcal{C}_r} \frac{1}{s!} \prod_{i=1}^s w(\ell_i; x).$$

Clearly, since \mathcal{C}_r is a finite set for every fixed r , this result is an immediate consequence of the following proposition:

Proposition 2.4. *For all $r > 0$,*

$$\sum_{s=1}^{\infty} \sum_{(\ell_1, \dots, \ell_s) \in \mathcal{C}_r \setminus \mathcal{C}_r^*} \frac{1}{s!} \prod_{i=1}^s w(\ell_i; x) = 0.$$

The proof of Proposition 2.4 is involved, and we postpone it to Section 2.3. For now, we assume that Proposition 2.4 and hence Theorem 2.3 hold, and explain how Theorems 1.1 and 1.2 follow from this.

Proof of Theorem 1.1. By splitting the sum over the set of loop configurations \mathcal{C}_r in Theorem 2.3 according to the lengths of the individual loops, using (1.6) we can write

$$\begin{aligned} Z(x) &= 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{1}{s!} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^s \left(\sum_{\ell \in \mathcal{L}_{r_i}} w(\ell; x) \right) \\ (2.3) \quad &= 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{1}{s!} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^s f_{r_i}(x). \end{aligned}$$

Now suppose that, given x , there exist $\gamma \in (0, 1)$ and $C < \infty$ such that

$$(2.4) \quad |f_r(x)| \leq C\gamma^r \quad \text{for all } r.$$

For future reference, we note that this condition is implied by the stronger condition that

$$(2.5) \quad \sum_{\ell \in \mathcal{L}_r} |w(\ell; x)| \leq C\gamma^r \quad \text{for all } r.$$

Under condition (2.4), if we write $h(r, s)$ for the summand in (2.3), we have

$$|h(r, s)| = \left| \frac{1}{s!} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^s f_{r_i}(x) \right| \leq \frac{C^s}{s!} \binom{r-1}{s-1} \gamma^r,$$

and thus

$$\sum_{s=1}^{\infty} \sum_{r=1}^{\infty} |h(r, s)| \leq \sum_{s=1}^{\infty} \frac{C^s}{s!} \sum_{r=s}^{\infty} \binom{r-1}{s-1} \gamma^r = \exp\left(\frac{C\gamma}{1-\gamma}\right) - 1.$$

Hence, we can apply Fubini's theorem to interchange the order of summation over r and s in (2.3), which yields

$$Z(x) = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \sum_{r=1}^{\infty} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^s f_{r_i}(x).$$

Note that under condition (2.4), $\sum_r f_r(x)$ is absolutely convergent. We now apply Mertens' theorem, which says that if a series $\sum_r a_r$ converges absolutely, and the series $\sum_r b_r$ converges, then their Cauchy product converges to $(\sum_r a_r)(\sum_r b_r)$. In particular, by induction, the s -fold Cauchy product of the series $\sum_r a_r$ with itself, which is $\sum_r \sum_{r_1+\dots+r_s=r} a_{r_1} a_{r_2} \dots a_{r_s}$, converges to $(\sum_r a_r)^s$. Applying this to $a_r = f_r(x)$, we obtain

$$(2.6) \quad Z(x) = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \left(\sum_{r=1}^{\infty} f_r(x) \right)^s = \exp\left(\sum_{r=1}^{\infty} f_r(x) \right).$$

Observe that this result holds already under the weaker of the two conditions (2.4) and (2.5), but that under the stronger condition (2.5), the loop

weights can in fact be summed in any order. We will now show that the condition of Theorem 1.1 implies (2.5). Indeed, under the condition of Theorem 1.1, there exists $\gamma \in (0, 1)$ such that $(d_{uv} - 1)|x_{uv}| \leq \gamma$ for all edges $uv \in E$. Observing that if a loop takes a step along uv , then there are at most $d_{uv} - 1$ possibilities for the next step, this implies that the sum of $|w(\ell; x)|$ over all loops ℓ of n steps is bounded by $|V|\gamma^n$. Since a loop of length r takes at least r steps, summing over $n \geq r$ yields (2.5). \square

Proof of Theorem 1.2. Recall definition (1.7) of the entries of $\Lambda(x)$, which are indexed by the directed representative edges of G . We can interpret this matrix as a transition matrix for non-backtracking paths on the graph G' which is represented by G . This represented graph G' can be obtained from G by removing every chain of additional edges from G , and identifying the two vertices at the ends of this chain (see Figure 2 for an example).

Indeed, consider two directed representative edges \vec{uv} and $\vec{wz} \neq \vec{vu}$ in G , and write \vec{uv}' and \vec{wz}' for the corresponding directed edges in the represented graph G' . By construction, a non-backtracking path in G' can make a step from \vec{uv}' to \vec{wz}' if and only if v is linked to w in the graph G , since only then will v be identified with w in G' . This step corresponds to either a direct step from \vec{uv} to \vec{wz} in G (if $v = w$), or to a sequence of steps along the chain linking v to w . In either case, the matrix entry $\Lambda_{\vec{uv}, \vec{wz}}(x)$ picks up all edge weights and turning angles associated with these steps in G .

We can now interpret this entry as describing the weight picked up by a non-backtracking walk in G' when it steps from \vec{uv}' to \vec{wz}' . Viewed in this way, the entry of the matrix $\Lambda^r(x)$ indexed by \vec{uv} and \vec{wz} is equal to the sum of the weights of all non-backtracking paths in G' of r steps starting from \vec{uv}' and ending on \vec{wz}' . In particular, the sum of the diagonal entries of $\Lambda^r(x)$ is equal to the sum of the weights of all non-backtracking paths in G' of r steps starting and ending on the same directed edge.

Now consider a loop ℓ of length r in G . Note that it is possible to start traversing this loop from each step it takes along a representative edge in two directions. Mapping the paths thus obtained to the represented graph G' yields precisely $2r/m(\ell)$ different non-backtracking paths of r steps in G' that start and end on the same directed edge. By (1.2), (1.5) and (1.6), it now follows that

$$\text{tr } \Lambda^r(x) = -2rf_r(x),$$

where the minus sign comes from the minus sign in the definition (1.5) of the sign of a loop in terms of its winding angle. Expressed in the eigenvalues $\lambda_i(x)$ of $\Lambda(x)$, we therefore have that

$$(2.7) \quad f_r(x) = -\frac{1}{2r} \sum_i \lambda_i^r(x).$$

In particular, condition (2.4) is satisfied if $\max_i |\lambda_i(x)| < 1$, so in this case the same argument as in the proof of Theorem 1.1 yields (2.6). Moreover,

if $\max_i |\lambda_i(x)| < 1$, then using (2.7) we can write

$$Z(x) = \exp\left(-\frac{1}{2} \sum_{r=1}^{\infty} \sum_i \frac{\lambda_i^r(x)}{r}\right) = \exp\left(-\frac{1}{2} \sum_i \sum_{r=1}^{\infty} \frac{\lambda_i^r(x)}{r}\right),$$

and since $\sum_{r=1}^{\infty} u^r/r = -\ln(1-u)$ if $|u| < 1$, we conclude that

$$Z(x) = \prod_i (1 - \lambda_i(x))^{1/2} = \sqrt{\det(\mathbf{I} - \Lambda(x))}. \quad \square$$

2.3. Cancellation of non-edge-disjoint loop configurations. We now turn to the missing step in the proofs of Theorems 1.1 and 1.2, which is the proof of Proposition 2.4. That is, we must show that the weights of all loop configurations (ℓ_1, \dots, ℓ_s) which are not edge-disjoint and satisfy $r(\ell_1) + \dots + r(\ell_s) = r$ for a given r , cancel each other. What complicates matters here, is the fact that these loop configurations do not cancel each other one by one, see for example Figure 5.¹ Our strategy of the proof is to map loop configurations to so-called *labelled* loop configurations, which do cancel each other one by one, and show that this implies cancellation of the unlabelled loop configurations for combinatorial reasons.

We will therefore start by introducing the notion of a *labelled* loop, and work our way from there towards the notion of a labelled loop configuration, and the proof of their cancellation. In words, a labelled loop is a loop with a label attached to each step it takes, where the labels are distinct positive integers. For periodic loops, the first step is repeated after completing a period, and we require that the label of the first step of the loop is smaller than the label associated with each of these repetitions.

Formally, a labelled loop ℓ^\diamond is a sequence $(v_0, a_0, v_1, a_1, \dots, v_{n-1}, a_{n-1})$ satisfying the following conditions:

- L1** $\ell = (v_0, \dots, v_{n-1})$ is a loop;
- L2** $(a_0, a_1, \dots, a_{n-1})$ is a sequence of distinct positive integers, called the *labelling* of the loop;
- L3** if ℓ is periodic, i.e. $m(\ell) > 1$, then a_0 is smaller than $a_{kn/m(\ell)}$ for all $k \in \{1, 2, \dots, m(\ell) - 1\}$.

We call the number a_i the *label on step $i + 1$* of the loop ℓ ; we also regard it as a label assigned to the edge $v_i v_{i+1}$. We will use the superscript \diamond for labelled loops, and the unlabelled loop corresponding to a labelled loop will consistently be denoted by dropping this superscript: if ℓ^\diamond is a labelled loop, then ℓ is the corresponding unlabelled loop, and so on.

Observe that one of the effects of labelling loops is that it breaks the periodicity of periodic loops: sequences representing labelled loops cannot be periodic. Therefore, if ℓ is periodic, we do not assign to the labelled loop

¹A picture of the same configurations appears in [6] to point out the error in Vdovichenko's paper; it is crucial here to take the multiplicity of the loops into account.

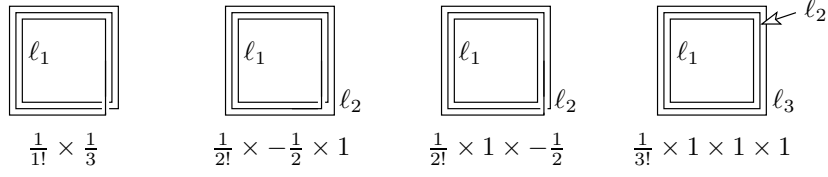


FIGURE 5. Four loop configurations on the same vertices and edges, where the traversals of the same edge have been drawn slightly apart to make them discernible. The factors $\frac{1}{s!} \prod_{i=1}^s \text{sgn}(\ell_i)/m(\ell_i)$ are spelled out below each loop configuration to show that the sum of their signed weights is 0.

$\ell^\diamond = (v_0, a_0, \dots, v_{n-1}, a_{n-1})$ the same weight as to its unlabelled counterpart. Instead, we define the weights of labelled loops in general by

$$(2.8) \quad w(\ell^\diamond; x) = \text{sgn}(\ell) \prod_{i=0}^{n-1} x_{v_i v_{i+1}},$$

where the sign is defined in terms of the winding angle of ℓ by (1.5), as before. Note that this weight is actually independent of the particular labelling of the loop, and that $w(\ell^\diamond; x) = m(\ell)w(\ell; x)$.

We write $n(\ell)$ for the number of steps of a loop ℓ (recall that this is not necessarily the same as the *length* $r(\ell)$ of the loop). By a *labelled loop configuration* we mean a collection $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ of labelled loops, in which all labels are distinct and take values from the set $\{1, 2, \dots, \sum_{i=1}^s n(\ell_i)\}$. In particular, any loop configuration (ℓ_1, \dots, ℓ_s) can be turned into a labelled loop configuration by attaching a label to every step of every loop in such a way, that condition **L3** above is fulfilled for every labelled loop obtained, and all labels $1, 2, \dots, \sum_{i=1}^s n(\ell_i)$ are used.

Now fix r and n , and consider a loop configuration (ℓ_1, \dots, ℓ_s) which is not edge-disjoint and satisfies $\sum_{i=1}^s r(\ell_i) = r$ and $\sum_{i=1}^s n(\ell_i) = n$. Let t denote the number of distinct loops in (ℓ_1, \dots, ℓ_s) , and write k_1, \dots, k_t for the respective number of times each of them occurs, so that $k_1 + \dots + k_t = s$. Consider the collection of all labelled loop configurations $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ that can be obtained from (ℓ_1, \dots, ℓ_s) by labelling the loops, as described above. For a periodic loop ℓ_i , only one of the rotations of its labelling, rotated over a multiple of the smallest period, satisfies condition **L3**. Furthermore, interchanging the labellings of two identical loops ℓ_i and ℓ_j yields the same labelled loop configuration. Therefore, the number of labelled loop configurations we obtain from (ℓ_1, \dots, ℓ_s) is precisely

$$\frac{n!}{\prod_{i=1}^s m(\ell_i) \prod_{i=1}^t k_i!}.$$

We assign to each of these labelled loop configurations the same weight $\prod_{i=1}^s w(\ell_i^\diamond; x)$, where we use the fact that according to the definition (2.8),

$w(\ell_i^\diamond; x)$ does not depend on the actual labelling. Then the total weight of all labelled loop configurations associated with (ℓ_1, \dots, ℓ_s) is

$$\frac{n!}{\prod_{i=1}^s m(\ell_i) \prod_{i=1}^t k_i!} \prod_{i=1}^s w(\ell_i^\diamond; x) = \frac{n!}{\prod_{i=1}^t k_i!} \prod_{i=1}^s w(\ell_i; x).$$

We claim that this is exactly $n!$ times the total weight that all the permutations of the loop configuration (ℓ_1, \dots, ℓ_s) contribute to the sum in Proposition 2.4. Indeed, there are precisely

$$\frac{s!}{\prod_{i=1}^s k_i!}$$

such permutations, and the weight each of them contributes to the sum is

$$\frac{1}{s!} \prod_{i=1}^s w(\ell_i; x).$$

We conclude that to prove Proposition 2.4, it suffices to show that for given n and r , the weights of all labelled loop configurations $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ such that $\sum_{i=1}^s n(\ell_i) = n$, $\sum_{i=1}^s r(\ell_i) = r$ and (ℓ_1, \dots, ℓ_s) is not edge-disjoint, sum to 0. Write $\mathcal{C}_{n,r}^\diamond$ for this collection of labelled loop configurations. We will now prove the desired cancellation of weights, and hence Proposition 2.4, by finding a bijection $g: \mathcal{C}_{n,r}^\diamond \rightarrow \mathcal{C}_{n,r}^\diamond$ which maps each labelled loop configuration to a labelled loop configuration which has a weight of the opposite sign, but with the same absolute value.

Proof of Proposition 2.4. Before we go into the formal details of the bijection, let us give an informal description of how it will work. Consider a labelled loop configuration $\{\ell_1^\diamond, \dots, \ell_s^\diamond\} \in \mathcal{C}_{n,r}^\diamond$, and let E^\diamond be the set of edges in G that are assigned more than 1 label in this configuration. Find the smallest of all the labels that are assigned to the edges in E^\diamond , let a be this label, and let uv be the edge to which this label is assigned. Next, find the second smallest label b which is assigned to the edge uv .

The label a labels a step of one of the loops ℓ_i . The label b either labels another step of the same loop ℓ_i , or it labels a step of a second loop ℓ_j , $i \neq j$. The bijection involves interchanging the “connections” on one side of the two steps marked a and b (either at the vertex u or at the vertex v), as illustrated in Figure 6. It is clear that this operation does not change the absolute value of the weight of the configuration, since the total number of steps that go through a given edge does not change. But Figure 6 also suggests that the operation corresponds to increasing or decreasing the number of “crossings” in the configuration by 1, which should indeed lead to a change in sign.

However, signs were formally defined in terms of winding angles, not numbers of crossings, since it is more difficult to make sense of the latter when loops are not edge-disjoint. Furthermore, we must still formally define the mapping g . We will now deal with these technical issues.

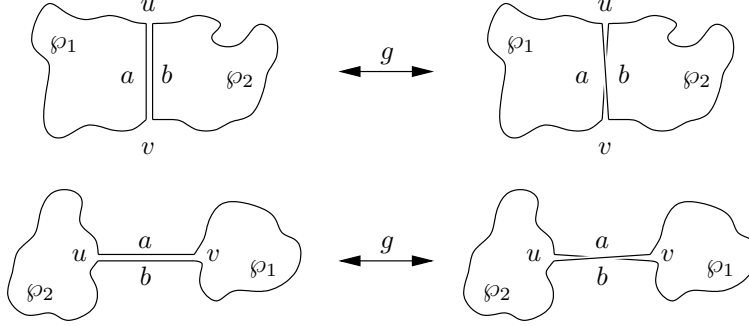


FIGURE 6. All cases that occur in the cancellation of labelled loop diagrams, as explained in the text. The curves \wp_1 and \wp_2 represent arbitrary paths connected to the vertices u and v .

For the formal treatment of the bijection, we need to introduce some additional notation. Given a sequence $a = (a_0, \dots, a_n)$ of arbitrary elements, we write a^{-1} for its *reversion* $a^{-1} = (a_n, a_{n-1}, \dots, a_0)$. If $b = (b_0, \dots, b_m)$ is another sequence of arbitrary elements, we write $a \oplus b$ for the concatenation of a with b , that is,

$$a \oplus b = (a_0, \dots, a_n, b_0, \dots, b_m).$$

The weight of a labelled loop configuration $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ is defined as the product of the signs of the loops ℓ_1, \dots, ℓ_s , times the product of all the edge weights picked up by all the loops. As was anticipated above, the product of edge weights will not change under the bijection, so we will only be concerned with the product of the signs of the loops. We recall from (1.4) and (1.5) that the sign of a loop $\ell = (v_0, \dots, v_{n-1})$ is defined in terms of its winding angle as

$$(2.9) \quad \text{sgn}(\ell) = -\exp\left(\frac{i}{2}\alpha(\ell)\right),$$

where the winding angle $\alpha(\ell)$ is given by

$$(2.10) \quad \alpha(\ell) = \sum_{i=0}^{n-1} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}).$$

We now define the winding angle and sign of a rooted loop (v_0, \dots, v_{n-1}) by the exact same formulas. In particular, all rotations of a loop ℓ have the same winding angle and sign. On the other hand, the reversion ℓ^{-1} of ℓ and all its rotations are traversed in the opposite direction, and therefore they all have winding angle $\alpha(\ell^{-1}) = -\alpha(\ell)$. However, since the winding angle of a loop is a multiple of 2π , we do have that

$$(2.11) \quad \text{sgn}(\ell^{-1}) = \text{sgn}(\ell) \quad \text{for all rooted loops } \ell.$$

We call all the rotations of a loop ℓ , and all rotations of its reversion ℓ^{-1} , alternative *representations* of ℓ . All these representations have the same sign. Likewise, the rotations of a labelled loop ℓ^\diamond and its reversion $\ell^{\diamond-1}$ will be called *representations* of this labelled loop.

We also need to define the winding angle for paths in G which are not loops. Recall that a path $\wp = (v_0, \dots, v_{n-1})$ is not a loop if $v_0 v_{n-1} \notin E$, $v_0 = v_{n-2}$, or $v_1 = v_{n-1}$. If we follow such a path from v_0 to v_{n-1} , we turn through $n - 2$ angles, and it is natural to define the winding angle of \wp by

$$\alpha(\wp) = \sum_{i=0}^{n-3} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}).$$

We now have all the notation we need to define and analyse the bijection formally. So consider a labelled loop configuration $\{\ell_1^\diamond, \dots, \ell_s^\diamond\} \in \mathcal{C}_{n,r}^\diamond$, and define E^\diamond , a , b and uv as above. We will now explain to which labelled loop configuration our configuration $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ is mapped by the bijection, and prove that the image has the opposite sign, and hence the opposite weight. There are three possible cases to consider, which are illustrated in Figure 6.

Case 1: *The labels a and b belong to different labelled loops.* Let ℓ_i^\diamond be the labelled loop containing label a , and let ℓ_j^\diamond be the labelled loop containing label b . Then these labelled loops have representations of the form $\hat{\ell}_i^\diamond = (u, a, v) \oplus \wp_1^\diamond$ and $\hat{\ell}_j^\diamond = (u, b, v) \oplus \wp_2^\diamond$, respectively, where \wp_1^\diamond and \wp_2^\diamond are paths interspersed with labels. We can now form the combined representation

$$\hat{\ell}_{ij}^\diamond = (u, a, v) \oplus \wp_1^\diamond \oplus (u, b, v) \oplus \wp_2^\diamond$$

of a new labelled loop ℓ_{ij}^\diamond . Our bijection maps $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ to the labelled loop configuration

$$\{\ell_1^\diamond, \dots, \ell_s^\diamond, \ell_{ij}^\diamond\} \setminus \{\ell_i^\diamond, \ell_j^\diamond\}.$$

To see that this labelled loop configuration has the opposite sign of its pre-image $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$, note that by (2.9)–(2.11),

$$\begin{aligned} \text{sgn}(\ell_i) \text{sgn}(\ell_j) &= \text{sgn}(\hat{\ell}_i) \text{sgn}(\hat{\ell}_j) = \exp\left(\frac{i}{2}\alpha(\hat{\ell}_i) + \frac{i}{2}\alpha(\hat{\ell}_j)\right) \\ &= \exp\left(\frac{i}{2}\alpha(\hat{\ell}_{ij})\right) = -\text{sgn}(\hat{\ell}_{ij}) = -\text{sgn}(\ell_{ij}). \end{aligned}$$

Case 2: *The labels a and b are on steps of the same labelled loop taken in the same direction.* This case is the reverse of Case 1. The labels a and b are in a labelled loop ℓ_i^\diamond which has a representation of the form

$$\hat{\ell}_i^\diamond = (u, a, v) \oplus \wp_1^\diamond \oplus (u, b, v) \oplus \wp_2^\diamond.$$

From this we obtain the representations $(u, a, v) \oplus \wp_1^\diamond$ and $(u, b, v) \oplus \wp_2^\diamond$ of two new labelled loops ℓ_{i1}^\diamond and ℓ_{i2}^\diamond . The bijection maps $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ to the labelled loop configuration

$$\{\ell_1^\diamond, \dots, \ell_s^\diamond, \ell_{i1}^\diamond, \ell_{i2}^\diamond\} \setminus \{\ell_i^\diamond\}.$$

The same argument as in Case 1 shows that $\text{sgn}(\ell_{i1})\text{sgn}(\ell_{i2}) = -\text{sgn}(\ell_i)$.

Case 3: *The labels a and b are on steps of the same labelled loop taken in opposite directions.* In this case the labels a and b are in a labelled loop ℓ_i^\diamond which has a representation of the form

$$\hat{\ell}_i^\diamond = (u, a, v) \oplus \wp_1^\diamond \oplus (v, b, u) \oplus \wp_2^\diamond.$$

From this we can construct the representation

$$\hat{\ell}^\diamond = (u, a, v) \oplus \wp_1^{\diamond-1} \oplus (v, b, u) \oplus \wp_2^\diamond$$

of a new labelled loop ℓ^\diamond . The bijection maps $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ to the labelled loop configuration

$$\{\ell_1^\diamond, \dots, \ell_s^\diamond, \ell^\diamond\} \setminus \{\ell_i^\diamond\}.$$

To verify that these loop configurations have opposite signs, observe that

$$(2.12) \quad \alpha(\hat{\ell}_i) = \alpha((u, v) \oplus \wp_1 \oplus (v, u)) + \alpha((v, u) \oplus \wp_2 \oplus (u, v)),$$

and likewise

$$(2.13) \quad \alpha(\hat{\ell}) = \alpha((u, v) \oplus \wp_1^{-1} \oplus (v, u)) + \alpha((v, u) \oplus \wp_2 \oplus (u, v)),$$

where \wp_1 and \wp_2 are the paths obtained from \wp_1^\diamond and \wp_2^\diamond by dropping the labels. Now notice that upon reversion,

$$(2.14) \quad \alpha((u, v) \oplus \wp_1 \oplus (v, u)) = -\alpha((u, v) \oplus \wp_1^{-1} \oplus (v, u)).$$

Furthermore, it is not difficult to see that

$$\alpha((u, v) \oplus \wp_1 \oplus (v, u)) = 2m\pi + \pi \quad \text{for some } m \in \mathbb{Z}.$$

Together with (2.12), (2.13) and (2.14), this implies

$$\frac{\text{sgn}(\ell_i)}{\text{sgn}(\ell)} = \frac{\text{sgn}(\hat{\ell}_i)}{\text{sgn}(\hat{\ell})} = \exp\left(\frac{i}{2}\alpha(\hat{\ell}_i) - \frac{i}{2}\alpha(\hat{\ell})\right) = -1.$$

We conclude that in all cases, the labelled loop configuration $\{\ell_1^\diamond, \dots, \ell_s^\diamond\}$ is mapped to a labelled loop configuration of opposite weight. From the explicit descriptions given above, it is not difficult to see that the mapping is bijective. As we have explained above, this implies Proposition 2.4. \square

3. PROOFS OF OUR RESULTS FOR THE ISING MODEL

In this section, we will apply Theorem 1.2 to the Ising model on the square lattice \mathbb{Z}^2 . We will see that this leads to explicit expressions for the free energy density and two-point functions in terms of sums over loops in \mathbb{Z}^2 or its dual \mathbb{Z}^{2*} , valid all the way up to the critical point. To get started, we must find a bound on the spectral radius of the matrix $\Lambda(x)$ appearing in Theorem 1.2. This will be done in Section 3.1. Then we will study the free energy density in Section 3.2, and finally the two-point functions at low and high temperatures in Sections 3.3 and 3.4, respectively.

3.1. Bound on the spectral radius. Let $G = (V, E)$ be a fixed finite rectangle in \mathbb{Z}^2 with no additional edges (i.e. $E_A = \emptyset$). Without loss of generality, we may assume that the vertex set is

$$(3.1) \quad V = \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\}.$$

Since we are on the square lattice, directed edges can point in only 4 directions, and we now introduce some convenient notation for this specific case. We write $v\uparrow$, $v\downarrow$, $v\rightarrow$ and $v\leftarrow$ for the directed edges from v to, respectively, $v+(0,1)$, $v-(0,1)$, $v+(1,0)$ and $v-(1,0)$. We also write $\uparrow v$ for the directed edge pointing from $v-(0,1)$ to v ; $\downarrow v$, $\rightarrow v$, $\leftarrow v$ are defined analogously.

Given a vector of edge weights $x = (x_{uv})_{uv \in E}$ on E , $\Lambda(x)$ is the transition matrix indexed by the directed edges of G , as defined in (1.7). For a vertex v not on the boundary of G , the row of $\Lambda(x)$ indexed by $\rightarrow v$, for instance, has exactly 3 nonzero entries, corresponding to the 3 possible steps that a loop can take from $\rightarrow v$. To be precise, with $u = v - (1, 0)$, these 3 entries are

$$\Lambda_{\rightarrow v, v\rightarrow}(x) = x_{uv}, \quad \Lambda_{\rightarrow v, v\uparrow}(x) = x_{uv}e^{i\pi/4}, \quad \Lambda_{\rightarrow v, v\downarrow}(x) = x_{uv}e^{-i\pi/4}.$$

Observe that most rows of $\Lambda(x)$ have exactly 3 nonzero entries. The only exceptions are the rows indexed by directed edges pointing to a vertex in ∂G . These exceptional rows make it impossible to compute the eigenvalues of $\Lambda(x)$ directly. We will therefore wrap G around a torus, so that all vertices can be treated alike, and then bound the eigenvalues of $\Lambda(x)$ in terms of those of the transition matrix on the torus, or equivalently, the transition matrix for a graph in the plane representing the torus.

To be precise, we first extend our graph G to a graph G^\odot (\odot stands for “torus”), by adding edges and vertices as shown in Figure 2 (left). Note that this adds directed representative edges $v\rightarrow$ and $\leftarrow v$ for every vertex v on the right boundary of G , and $v\uparrow$ and $\downarrow v$ for v on the top boundary. All other edges that are added are considered as *additional* edges in the graph G^\odot . Henceforth, when we work on the graph G^\odot , computations will be performed modulo M and N in the two respective lattice directions.

The matrix Λ^\odot is the transition matrix for the graph G^\odot , with specific edge weights chosen as follows: all representative edges of G^\odot have edge weight 1; for the additional edges, we choose the edge weights in such a way, that the product of the edge weights along every chain of additional edges linking opposite sides of the rectangle to each other, is -1 . Note that by this choice, the factor -1 will exactly compensate the sign picked up by a path which follows the chain, because of the 4 quarter-turns it makes.

Proof of Theorem 1.3. We first prove that the spectral radius of the matrix Λ^\odot is $\sqrt{2} + 1$. To this end, assume that the rows of Λ^\odot are arranged in such a way, that for every vertex $v \in V$, the 4 rows indexed by $\rightarrow v$, $\uparrow v$, $\leftarrow v$ and $\downarrow v$ immediately succeed each other in this order. Let Π be the permutation matrix which permutes the columns of Λ^\odot so that column vd maps to column dv , for all $v \in V$ and $d \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$.

By construction, the matrix $\Lambda^\odot \Pi$ with the permuted columns is now a block-diagonal matrix, since the 4 rows indexed by the directed edges pointing to v are matched along the diagonal with the 4 columns indexed by the directed edges pointing out from v . By considering the turning angles, it is easy to see that each 4×4 block is equal to the Hermitian matrix

$$A = \begin{bmatrix} 1 & \exp(i\pi/4) & 0 & \exp(-i\pi/4) \\ \exp(-i\pi/4) & 1 & \exp(i\pi/4) & 0 \\ 0 & \exp(-i\pi/4) & 1 & \exp(i\pi/4) \\ \exp(i\pi/4) & 0 & \exp(-i\pi/4) & 1 \end{bmatrix},$$

which has eigenvalues $\sqrt{2} + 1$ and $\sqrt{2} - 1$, both of multiplicity 2.

For a square matrix B of dimensions d , let $\|B\|$ denote the operator norm induced by the Euclidean norm on \mathbb{C}^d . Since A is Hermitian, its spectral radius is equal to $\|A\|$. It follows that the operator norm of $\Lambda^\odot \Pi$ is given by $\|A\| = \sqrt{2} + 1$, and since permuting columns does not change the operator norm of a matrix, we conclude that $\|\Lambda^\odot\| = \sqrt{2} + 1$.

In general, the spectral radius of a matrix B is bounded from above by the operator norm $\|B\|$. We will now use this fact, together with the submultiplicativity of the operator norm, to bound the spectral radius of the matrix $\Lambda(x)$. To this end, let $D(x)$ be the diagonal matrix of the same dimensions as Λ^\odot , defined as follows. For vertices v on the right boundary of G , the diagonal entries of $D(x)$ on the rows $v \rightarrow$ and $\leftarrow v$ are 0, and so are the diagonal entries on the rows $v \uparrow$ and $\downarrow v$ for v on the top boundary of G . For all other directed edges \vec{uv} in the graph G^\odot , the diagonal entry of $D(x)$ on row \vec{uv} is equal to the edge weight x_{uv} .

Now consider the matrix $D(x)\Lambda^\odot$. This is just the matrix Λ^\odot with all rows corresponding to directed edges \vec{uv} in the graph G multiplied by x_{uv} , and all rows corresponding to directed edges which are in the graph G^\odot , but not in the graph G , zeroed out. Therefore, considered as a transition matrix on the graph G^\odot , $D(x)\Lambda^\odot$ generates exactly the same loops as the matrix $\Lambda(x)$, with the exact same weights.

Let ψ be an eigenvector of $\Lambda(x)$, indexed by the directed edges in G . Extend this vector to a vector ψ^\odot indexed by the directed representative edges in G^\odot , by setting $\psi_{\vec{uv}}^\odot = \psi_{\vec{uv}}$ if uv is an edge in G and $\psi_{\vec{uv}}^\odot = 0$ if uv is a representative edge in G^\odot but not in G . Then, by construction, ψ^\odot is an eigenvector of $D(x)\Lambda^\odot$ with the same eigenvalue as ψ . Hence all eigenvalues of $\Lambda(x)$ are eigenvalues of $D(x)\Lambda^\odot$, and we conclude that

$$\max_i |\lambda_i(x)| \leq \|D(x)\Lambda^\odot\| \leq \|D(x)\| \cdot \|\Lambda^\odot\| = (\sqrt{2} + 1)\|x\|_\infty.$$

The desired bound on $|f_r(x)|$ now follows from (2.7) and the fact that the number of directed edges in G is bounded by $4|V|$. \square

Remark. Using Fourier transforms, we can actually compute all eigenvalues of Λ^\odot . Also, with some extra effort, it is possible to show that for all finite rectangles, the spectral radius of $\Lambda(x)$ is *strictly* less than $(\sqrt{2} + 1)\|x\|_\infty$.

3.2. Free energy density. We are now going to use the bound obtained in Theorem 1.3 to derive an explicit expression for $f(\beta)$ as a power series in the parameter $\exp(-2\beta)$ for $\beta > \beta_c$, and in the parameter $\tanh \beta$ for $\beta < \beta_c$. From this expression we then also obtain Onsager's formula.

The coefficients of the power series for $f(\beta)$ are determined by the loops in \mathbb{Z}^2 for which the origin $o = (0, 0)$ is the lexicographically smallest vertex they traverse. To be more specific, we define \mathcal{L}_r° as the collection of all loops $\ell = (v_0, \dots, v_{r-1})$ in \mathbb{Z}^2 such that $v_0 = o$. Here, all edges in \mathbb{Z}^2 are regarded as representative edges, so that the length r of a loop is equal to its number of steps. Let all edges of \mathbb{Z}^2 have the same edge weight x . Then we can define the weights $w(\ell; x)$ of all loops $\ell \in \mathcal{L}_r^\circ$ by (1.2), as before (by slight abuse of notation, x now stands both for the weight of a single edge, and for the vector of all edge weights). Write

$$f_r^\circ(x) = \sum_{\ell \in \mathcal{L}_r^\circ} w(\ell; x).$$

Observe that $f_r^\circ(x) = f_r^\circ(1)x^r$, so $\sum_r f_r^\circ(x)$ is really a power series in x .

Theorem 3.1. *For all $\beta \in (0, \beta_c) \cup (\beta_c, \infty)$, the free energy density satisfies*

$$(3.2) \quad -\beta f(\beta) - \ln(2 \cosh 2\beta) = -\ln(1 + x^2) + \sum_{r=1}^{\infty} f_r^\circ(x),$$

where $x = \exp(-2\beta)$ for $\beta \in (\beta_c, \infty)$ and $x = \tanh \beta$ for $\beta \in (0, \beta_c)$.

Proof. We start with the high-temperature case, so fix $\beta \in (0, \beta_c)$ and set $x = \tanh \beta$. Let G be a rectangle in \mathbb{Z}^2 , and take the set of additional edges to be empty. Note that by (1.14), $x \in (0, \sqrt{2} - 1)$. By (1.13), we have

$$(3.3) \quad \ln Z_{G,\beta}^{\text{free}} = |V| \ln 2 + |E| \ln(\cosh \beta) + \ln Z_G(x),$$

where $Z_G(x)$ is the generating function for the graph G with edge weights equal to x . By Theorems 1.2 and 1.3, $\ln Z_G(x)$ is equal to $\sum_r f_{G,r}(x)$, where $f_{G,r}(x)$ denotes the sum of the weights of all loops of length r in the graph G .

Consider these loops of length r in G . For each vertex $v \in V$, let $\mathcal{L}_r^v(G)$ denote the collection of loops in G of length r for which v is the smallest vertex traversed. Observe that if v has distance at least r to the boundary of G , then $\mathcal{L}_r^v(G)$ can be mapped bijectively to \mathcal{L}_r° by a translation on \mathbb{Z}^2 , hence $\sum_{\ell \in \mathcal{L}_r^v(G)} w(\ell; x) = f_r^\circ(x)$. There are at most $|\partial G|r$ vertices at a distance less than r from ∂G , and since $|x| < 1$, for such a vertex v we have $\sum_{\ell \in \mathcal{L}_r^v(G)} |w(\ell; x)| \leq 3^r$ (by counting non-backtracking paths). From these observations and the fact that $|\partial G|/|V| \rightarrow 0$ as $G \rightarrow \mathbb{Z}^2$, it follows that

$$\lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} f_{G,r}(x) = \lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \sum_{v \in V} \sum_{\ell \in \mathcal{L}_r^v(G)} w(\ell; x) = f_r^\circ(x) \quad \text{for all } r \geq 1.$$

Furthermore, Theorem 1.3 says that $|f_{G,r}(x)| \leq 2|V|(\sqrt{2}+1)^r x^r$. Therefore, by dominated convergence and Theorem 1.2,

$$\lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \ln Z_G(x) = \lim_{G \rightarrow \mathbb{Z}^2} \sum_{r=1}^{\infty} \frac{1}{|V|} f_{G,r}(x) = \sum_{r=1}^{\infty} f_r^{\circ}(x).$$

We now combine this with (3.3), and use $\lim_{G \rightarrow \mathbb{Z}^2} |E|/|V| = 2$ to obtain

$$-\beta f(\beta) = \lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \ln Z_{G,\beta}^{\text{free}} = \ln(2 \cosh^2 \beta) + \sum_{r=1}^{\infty} f_r^{\circ}(x).$$

Using $\cosh^2 \beta + \sinh^2 \beta = \cosh^2 \beta (1 + x^2) = \cosh 2\beta$, we can cast this into the form of equation (3.2). We perform this final step here to obtain a common expression for $f(\beta)$ for all non-critical temperatures. Indeed, the low-temperature case can be treated in a similar manner, except that one must work on the dual graphs G^* with edge weights $x = \exp(-2\beta)$ on the dual edges, and use (1.12) instead of (1.13). In this case one obtains

$$-\beta f(\beta) = \lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \ln Z_{G,\beta}^+ = 2\beta + \sum_{r=1}^{\infty} f_r^{\circ}(x),$$

which can also be cast into the form of equation (3.2), by using the logarithm of the equality $2 \cosh 2\beta = e^{2\beta}(1 + x^2)$ to rewrite the 2β term. \square

Proof of Theorem 1.4. The expression for $f(\beta)$ derived above establishes directly that the free energy density is analytic on $(0, \beta_c) \cup (\beta_c, \infty)$. We will now show that $f(\beta)$ is given by Onsager's formula.

In the proof of Theorem 3.1, we have obtained $\sum_r f_r^{\circ}(x)$ as the limit of $|V|^{-1} \sum_r f_{G,r}(x)$. It is clear from the proof that here we may as well replace $f_{G,r}(x)$ by the corresponding sum of loop weights for the periodic graph G^{\odot} from Section 3.1. In fact, the argument becomes even simpler on G^{\odot} , since we no longer have to treat vertices near the boundary separately. The transition matrix generating the loops in G^{\odot} with the desired edge weights x is $x\Lambda^{\odot}$. Hence, by Theorem 1.2 and the proof of Theorem 1.3 in Section 3.1, we see that for $x \in (0, \sqrt{2} - 1)$,

$$(3.4) \quad \sum_{r=1}^{\infty} f_r^{\circ}(x) = \lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \sum_{r=1}^{\infty} f_{G^{\odot},r}(x) = \lim_{G \rightarrow \mathbb{Z}^2} \frac{1}{|V|} \frac{1}{2} \ln \det(\mathbf{I} - x\Lambda^{\odot}).$$

We can compute $\det(\mathbf{I} - x\Lambda^{\odot})$ by taking the Fourier transform of Λ^{\odot} , a computation which has appeared before, for instance in [9, 21, 25]. We give the calculation here for completeness, but will go through it a bit quickly.

Without loss of generality, we may assume that V is the set (3.1), in which case the Fourier transform of Λ^{\odot} is defined as

$$\tilde{\Lambda}_{(p,q)d,(p',q')d'}^{\odot} = \frac{1}{MN} \sum_{k,k'=0}^{M-1} \sum_{l,l'=0}^{N-1} e^{-\frac{2\pi i}{M}(pk-p'k') - \frac{2\pi i}{N}(ql-q'l')} \Lambda_{(k,l)d,(k',l')d'}^{\odot},$$

where $d, d' \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$. The calculation of this Fourier transform is made straightforward by the periodicity of Λ^\odot . The calculation reveals that the only entries surviving the summations are those for which $p' = p$ and $q' = q$. Hence, $\tilde{\Lambda}^\odot$ is a block-diagonal matrix of 4×4 blocks. To be precise, writing $\omega_p = 2\pi p/M$, $\omega_q = 2\pi q/N$, the 4×4 block for given p and q is

$$\tilde{\Lambda}_{(p,q) \cdot, (p,q) \cdot}^\odot = \begin{bmatrix} e^{i\omega_p} & e^{i\omega_p + i\pi/4} & 0 & e^{i\omega_p - i\pi/4} \\ e^{i\omega_q - i\pi/4} & e^{i\omega_q} & e^{i\omega_q + i\pi/4} & 0 \\ 0 & e^{-i\omega_p - i\pi/4} & e^{-i\omega_p} & e^{-i\omega_p + i\pi/4} \\ e^{-i\omega_q + i\pi/4} & 0 & e^{-i\omega_q - i\pi/4} & e^{-i\omega_q} \end{bmatrix}.$$

Since $\det(\mathbf{I} - x\Lambda^\odot) = \det(\mathbf{I} - x\tilde{\Lambda}^\odot)$, from this Fourier transform we obtain

$$\begin{aligned} \det(\mathbf{I} - x\Lambda^\odot) &= \prod_{p=0}^{M-1} \prod_{q=0}^{N-1} \det(\mathbf{I} - x\tilde{\Lambda}_{(p,q) \cdot, (p,q) \cdot}^\odot) \\ &= \prod_{p=0}^{M-1} \prod_{q=0}^{N-1} [(1+x^2)^2 - 2x(1-x^2)(\cos \omega_p + \cos \omega_q)]. \end{aligned}$$

Using (3.4), we conclude that

$$\begin{aligned} (3.5) \quad \sum_{r=1}^{\infty} f_r^\odot(x) &= \lim_{M, N \rightarrow \infty} \frac{1}{2MN} \ln \det(\mathbf{I} - x\Lambda^\odot) \\ &= \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln[(1+x^2)^2 - 2x(1-x^2)(\cos \omega_1 + \cos \omega_2)] d\omega_1 d\omega_2. \end{aligned}$$

To finish the computation, note that by (3.2), we have

$$\begin{aligned} -\beta f(\beta) &= \ln \left[\frac{2 \cosh 2\beta}{1+x^2} \right] + \sum_{r=1}^{\infty} f_r^\odot(x) \\ &= \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln \left[\frac{4 \cosh^2 2\beta}{(1+x^2)^2} \right] d\omega_1 d\omega_2 + \sum_{r=1}^{\infty} f_r^\odot(x). \end{aligned}$$

Combining this with (3.5), and then using the identity

$$\frac{2x(1-x^2)}{(1+x^2)^2} = \frac{\sinh 2\beta}{\cosh^2 2\beta},$$

which holds both for $x = \exp(-2\beta)$ and for $x = \tanh \beta$, we obtain

$$-\beta f(\beta) = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln[4 \cosh^2 2\beta - 4 \sinh 2\beta (\cos \omega_1 + \cos \omega_2)] d\omega_1 d\omega_2.$$

This is Onsager's formula for the isotropic Ising model on \mathbb{Z}^2 . \square

3.3. Low-temperature correlations. In this section, we discuss the Ising model with positive boundary conditions. We consider rectangles $G = (V, E)$ in \mathbb{Z}^2 (which later tend to \mathbb{Z}^2) and denote by G^* the weak dual of G . Recall that every spin configuration $\sigma \in \Omega^+$ on G corresponds bijectively to an even subgraph of G^* , that is, a graph in which all vertices in V^* have even degree. For given σ , we denote the corresponding even subset of E^* by $F(\sigma)$; for given even $F \subset E^*$, we denote the corresponding spin configuration by $\sigma(F)$.

Setting $x_e = e^{-2\beta}$ for every edge e in \mathbb{Z}^{2*} , by (1.11) and (1.12) we have

$$(3.6) \quad P_{G,\beta}^+(\sigma) = \frac{1}{Z_{G^*}(x)} \prod_{e \in F(\sigma)} x_e, \quad \sigma \in \Omega^+,$$

where $Z_{G^*}(x)$ is the generating function for G^* with edge weight vector $x = (x_e)_{e \in E^*}$. Note that here we have defined the edge weights on all edges of \mathbb{Z}^{2*} , which is necessary because our graphs G will tend to \mathbb{Z}^2 . However, the weight vector x is implicitly restricted to the relevant edges when we work on a given graph G . Such implicit restrictions to the relevant edges will occur throughout this and the following section.

Proof of Theorem 1.5. Fix $u, v \in \mathbb{Z}^2$, $u \neq v$, and let γ be a self-avoiding path in \mathbb{Z}^2 from u to v . We may assume that G is large enough so that u , v and γ are all contained in the area spanned by G , see Figure 7. We will express the two-point function $\langle \sigma_u \sigma_v \rangle_{G,\beta}^+$ as the quotient of two generating functions. To this end, we define new edge weights x'_e on the edges of \mathbb{Z}^{2*} such that $x'_e = -x_e$ if e crosses γ , and $x'_e = x_e$ otherwise. The reason for defining the weights x'_e in this way is the crucial fact that for all $\sigma \in \Omega^+$,

$$(3.7) \quad \sigma_u \sigma_v \prod_{e \in F(\sigma)} x_e = \prod_{e \in F(\sigma)} x'_e.$$

To see this, recall that the edges in $F(\sigma)$, by their very definition, cross edges $xy \in E$ for which $\sigma_x \neq \sigma_y$. If $\sigma_u = \sigma_v$, then following γ from u to v , we necessarily cross an even number of such edges. If $\sigma_u \neq \sigma_v$, then we cross an odd number. In either case, (3.7) holds.

With the help of (3.6), we can write

$$\langle \sigma_u \sigma_v \rangle_{G,\beta}^+ = \sum_{\sigma \in \Omega^+} \sigma_u \sigma_v P_{G,\beta}^+(\sigma) = \frac{1}{Z_{G^*}(x)} \sum_{\sigma \in \Omega^+} \sigma_u \sigma_v \prod_{e \in F(\sigma)} x_e,$$

and using (3.7) and the bijection between even F and Ω^+ , we obtain

$$(3.8) \quad \langle \sigma_u \sigma_v \rangle_{G,\beta}^+ = \frac{1}{Z_{G^*}(x)} \sum_{\text{even } F \subset E^*} \prod_{e \in F} x'_e = \frac{Z_{G^*}(x')}{Z_{G^*}(x)}.$$

The idea that correlations in the Ising model can be studied by means of ratios of generating functions with changed edge weights (or equivalently,

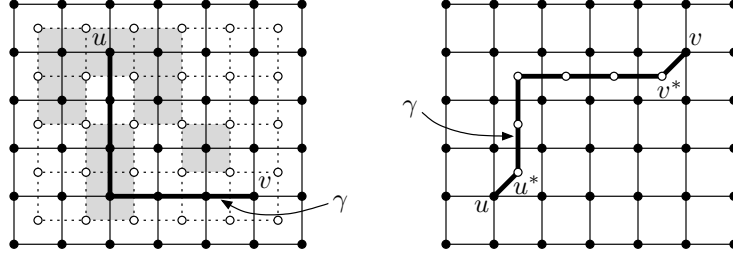


FIGURE 7. The paths γ (with bold edges), that we use to study the 2-point functions $\langle \sigma_u \sigma_v \rangle_{G, \beta}^{\square}$. The low-temperature case is on the left (spins are -1 in the gray squares, $+1$ in the white regions), the high-temperature case on the right.

changed spin-spin interactions) has arisen before in the physics literature, see [13]. It now follows from Theorems 1.2 and 1.3 that for $\beta > \beta_c$, we have

$$\langle \sigma_u \sigma_v \rangle_{G, \beta}^+ = \exp \left(\sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}_r(G^*)} [w(\ell; x') - w(\ell; x)] \right),$$

where $\mathcal{L}_r(G^*)$ is the collection of loops of length r in the graph G^* .

Call a loop in G^* *uv-odd* if it crosses γ an odd number of times. It is not difficult to see that this notion depends only on u and v , and not on the particular choice of γ . Observe that for *uv-odd* loops ℓ , $w(\ell; x') = -w(\ell; x)$, while for loops ℓ that are not *uv-odd*, $w(\ell; x') = w(\ell; x)$. It follows that

$$(3.9) \quad \langle \sigma_u \sigma_v \rangle_{G, \beta}^+ = \exp \left(-2 \sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(G^*): \\ \ell \text{ uv-odd}}} w(\ell; x) \right).$$

Note that a *uv-odd* loop of length r cannot travel far from u and v . To be precise, these loops must be contained in $B_r^u \cup B_r^v$, where B_r^u is a square in the plane of side length r centred at u , and B_r^v is defined similarly. To study the convergence of (3.9) as $G \rightarrow \mathbb{Z}^2$, for arbitrary rectangles R in \mathbb{Z}^2 that can be finite or infinite, and even equal to \mathbb{Z}^2 , we now define

$$a_r(R^*; x) := \sum_{\substack{\ell \in \mathcal{L}_r(R^*): \\ \ell \text{ uv-odd}}} w(\ell; x).$$

This definition makes sense both for finite and infinite R , since the loops contributing to the sum must be contained in $B_r^u \cup B_r^v$.

Let B_r^{uv} denote the smallest rectangle in \mathbb{R}^2 containing both B_r^u and B_r^v , and write $R^* \cap B_r^{uv}$ for the largest subgraph of R^* which is a rectangle in \mathbb{Z}^{2*} entirely contained in B_r^{uv} . Then for all R ,

$$(3.10) \quad a_r(R^*; x) = a_r(R^* \cap B_r^{uv}; x) = \frac{1}{2} \sum_{\ell \in \mathcal{L}_r(R^* \cap B_r^{uv})} [w(\ell; x) - w(\ell; x')].$$

Now, since the volume of B_r^{uv} is bounded from above by $(\|u - v\| + r)^2$, and $\exp(2\beta_c) = \sqrt{2} + 1$ by (1.14), Theorem 1.3 yields the uniform bound

$$(3.11) \quad |a_r(R^*; x)| \leq 2(\|u - v\| + r)^2 \exp(-2(\beta - \beta_c)r) \quad \text{for all } R.$$

We now return to (3.9). Since eventually, $G^* \cap B_r^{uv} = \mathbb{Z}^{2*} \cap B_r^{uv}$ when $G \rightarrow \mathbb{Z}^2$, from (3.10) we conclude that

$$a_r(G^*; x) \rightarrow a_r(\mathbb{Z}^{2*}; x) \quad \text{for all } r \geq 1.$$

Moreover, the $a_r(G^*; x)$ are uniformly bounded in G by the right-hand side of (3.11), which is summable over r . Therefore, by dominated convergence,

$$\lim_{G \rightarrow \mathbb{Z}^2} \sum_{r=1}^{\infty} a_r(G^*; x) = \sum_{r=1}^{\infty} a_r(\mathbb{Z}^{2*}; x),$$

where the series on the right is absolutely summable. Using (3.9), this gives the following result, which we state here as an intermediate theorem because of its significance, and for future reference:

Theorem 3.2. *For all $\beta \in (\beta_c, \infty)$ and fixed $u, v \in \mathbb{Z}^2$ ($u \neq v$),*

$$\lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle_{G, \beta}^+ = \exp \left(-2 \sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(\mathbb{Z}^{2*}): \\ \ell \text{ } uv\text{-odd}}} w(\ell; x) \right) =: \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^+ > 0.$$

Next, we consider $\langle \sigma_u \rangle_{G, \beta}^+$ for $u \in G \setminus \partial G$. We can treat this like $\langle \sigma_u \sigma_v \rangle_{G, \beta}^+$ by taking v on the boundary of G , since then $\sigma_v = +1$. We now call a loop which crosses γ an odd number of times u -odd, since this notion depends only on u , not on v . The box B_r^{uv} can be replaced by B_r^u in the argument, which replaces $(\|u - v\| + r)^2$ by r^2 in (3.11). In this case, the result is:

Theorem 3.3. *For all $\beta \in (\beta_c, \infty)$ and fixed $u \in \mathbb{Z}^2$,*

$$\lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \rangle_{G, \beta}^+ = \exp \left(-2 \sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(\mathbb{Z}^{2*}): \\ \ell \text{ } u\text{-odd}}} w(\ell; x) \right) =: \langle \sigma_u \rangle_{\mathbb{Z}^2, \beta}^+ > 0.$$

The limit in Theorem 3.3 is easily seen to be independent of the choice of u , and we take $u = o$ as the canonical choice. We now consider what happens to the two-point function when we take u and v further and further apart. When $r < \|u - v\|/2$, the boxes B_r^u and B_r^v are disjoint. If this is the case, a uv -odd loop of length r in \mathbb{Z}^{2*} must be contained in either B_r^u or B_r^v . Hence we have

$$\begin{aligned} \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^+ &= \exp \left(-2 \sum_{r \geq \|u - v\|/2} a_r(\mathbb{Z}^{2*}; x) \right) \\ &\quad \times \exp \left(-2 \sum_{r < \|u - v\|/2} [a_r(\mathbb{Z}^{2*} \cap B_r^u; x) + a_r(\mathbb{Z}^{2*} \cap B_r^v; x)] \right). \end{aligned}$$

When $\|u - v\| \rightarrow \infty$, the first factor converges to 1 exponentially fast, since the uniform bound in (3.11) applies to $a_r(\mathbb{Z}^{2*}; x)$. In the second factor, the first term in the sum is a sum over the u -odd loops of length r , and the second term is a sum over the v -odd loops. Hence the second factor factorizes and converges (exponentially fast) to $[\langle \sigma_o \rangle_{\mathbb{Z}^2, \beta}^+]^2$. \square

3.4. High-temperature correlations. In this section, we discuss the Ising model on rectangles $G = (V, E)$ in \mathbb{Z}^2 (which will again tend to \mathbb{Z}^2) with free boundary conditions. Our aim is to study the two-point functions $\langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}}$ at high temperatures. From the definitions (1.9), (1.10) and (1.15) we have

$$\langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \sum_{\sigma \in \Omega^{\text{free}}} \sigma_u \sigma_v P_{G, \beta}^{\text{free}}(\sigma) = \frac{1}{Z_{G, \beta}^{\text{free}}} \sum_{\sigma \in \Omega^{\text{free}}} \sigma_u \sigma_v \prod_{xy \in E} e^{\beta \sigma_x \sigma_y}.$$

Performing the high-temperature expansion on the right-hand side of this expression, in the way explained in Section 1.2.1, leads to

$$\langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \frac{2^{|V|} (\cosh \beta)^{|E|}}{Z_{G, \beta}^{\text{free}}} \sum_{F \subset E: \delta F = \{u, v\}} \prod_{e \in F} x_e,$$

where we set $x_e = \tanh \beta$ for every edge e in \mathbb{Z}^2 , and δF denotes the set of all vertices that have odd degree in (V, F) . Using (1.13), we conclude that

$$(3.12) \quad \langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \frac{1}{Z_G(x)} \sum_{F \subset E: \delta F = \{u, v\}} \prod_{e \in F} x_e,$$

where $Z_G(x)$ is the graph generating function for the graph G with edge weight vector $x = (x_e)_{e \in E}$.

Proof of Theorem 1.6. As in the low-temperature case, we first fix $u, v \in \mathbb{Z}^2$, $u \neq v$, and we will again work with a path connecting u with v . However, this time, we will also use this path to construct a modified graph, in which the edges along the path are included as additional edges.

To be precise, we first choose dual vertices u^* and v^* of \mathbb{Z}^{2*} such that $\|u - u^*\| = \|v - v^*\| = 1/2$, and then we pick a self-avoiding path γ in \mathbb{Z}^{2*} from u^* to v^* . Let V_γ denote the set of vertices in γ , and let E_γ denote the set of edges traversed by γ , together with the edges uu^* and vv^* . For an arbitrary rectangle R in \mathbb{Z}^2 (either finite or infinite) containing u and v , we denote by R_γ the graph obtained from R by adding all vertices in V_γ to its vertex set, and all edges in E_γ to its edge set. In R_γ , all edges added from the set E_γ are considered as additional edges, and the edges from R are considered as representative. See also Figure 7 (right). We can in particular augment \mathbb{Z}^2 in the way described above to obtain the graph \mathbb{Z}_γ^2 .

As in the low-temperature case, we now define weights x'_e on the edge set of \mathbb{Z}^2 such that $x'_e = -x_e$ if e crosses γ , and $x'_e = x_e$ otherwise. We also

define edge weights $x'_\gamma(t)_e$ on the edge set of \mathbb{Z}_γ^2 , as follows:

$$x'_\gamma(t)_e = \begin{cases} x'_e & \text{if } e \text{ is an edge of } \mathbb{Z}^2; \\ 1 & \text{if } e \in E_\gamma \setminus \{uu^*\}; \\ t & \text{if } e = uu^*. \end{cases}$$

To motivate this definition, consider a given rectangle $G = (V, E)$ in \mathbb{Z}^2 , large enough so that $u, v \in V$. We claim that

$$(3.13) \quad \sum_{\substack{F \subseteq E: \\ \delta F = \{u, v\}}} \prod_{e \in F} x_e = \sum_{\substack{\text{even } F \subseteq E \cup E_\gamma: \\ F \supset E_\gamma}} (-1)^{C_F} \prod_{e \in F} x'_\gamma(1)_e.$$

To see this, note that we can bijectively map every F contributing to the first sum to a subgraph in the second sum, by taking the union $F \cup E_\gamma$. Doing this may introduce edge crossings, whence the factor $(-1)^{C_F}$, but these are compensated by switching from the edge weight vector x to $x'_\gamma(1)$.

The crucial step is now to recognize the last expression as the derivative of a graph generating function. Indeed, a simple consideration shows that

$$\sum_{\substack{\text{even } F \subseteq E \cup E_\gamma: \\ F \supset E_\gamma}} (-1)^{C_F} \prod_{e \in F} x'_\gamma(1)_e = \frac{\partial}{\partial t} \left(\sum_{\substack{\text{even } F \subseteq E \cup E_\gamma}} (-1)^{C_F} \prod_{e \in F} x'_\gamma(t)_e \right),$$

evaluated at any t , since any even $F \subset E_\gamma$ contributes at most one factor t to the product of edge weights on the right. In particular, we are allowed to evaluate the derivative at $t = 0$. By (3.12) and (3.13), this establishes that

$$(3.14) \quad Z_G(x) \cdot \langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \frac{\partial}{\partial t} Z_{G_\gamma}(x'_\gamma(t)) \Big|_{t=0}.$$

We now fix $\beta \in (0, \beta_c)$, so that by (1.14), $x_e \in (0, \sqrt{2} - 1)$ for every $e \in E$, and by Theorem 1.3, the spectral radius of $\Lambda(x')$ is strictly less than 1. We now need a similar bound on the spectral radius of the matrix $\Lambda_\gamma(x'_\gamma(t))$, which is the transition matrix for the modified graph G_γ with edge weight vector $x'_\gamma(t)$. The difference between these two matrices is that $\Lambda_\gamma(x'_\gamma(t))$ allows transitions between u and v along the chain of additional edges in E_γ . This means that the 32 matrix entries from du to vd' and from dv to ud' , with $d, d' \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$, are nonzero in $\Lambda_\gamma(x'_\gamma(t))$ for $t \neq 0$, while they are 0 in $\Lambda(x')$; all other entries of the two matrices are the same.

The 32 deviating matrix entries are all of the form $te^{i\phi/2}$, where ϕ is a sum of turning angles. Here, t will be treated as a complex variable. For $t = 0$, $\Lambda_\gamma(x'_\gamma(t)) = \Lambda(x')$. Since the eigenvalues vary continuously with t , we conclude that there exists $\varepsilon > 0$ such that for all t satisfying $|t| < \varepsilon$, the spectral radius of $\Lambda_\gamma(x'_\gamma(t))$ is bounded from above by some $\alpha \in (0, 1)$.

Hence, if $|t| < \varepsilon$, Theorem 1.2 applies, and we obtain

$$Z_{G_\gamma}(x'_\gamma(t)) = \exp \left(\sum_{r=1}^{\infty} f_{\gamma r}(t) \right),$$

where

$$f_{\gamma r}(t) = \sum_{\ell \in \mathcal{L}_r(G_\gamma)} w(\ell; x'_\gamma(t)).$$

Note that, this last sum being finite, the $f_{\gamma r}(t)$ are polynomials in t . Also, from (2.7) it follows that $|f_{\gamma r}(t)| \leq 2|V|\alpha^r$. Therefore, the partial sums of the series $\sum_r f_{\gamma r}(t)$ are uniformly convergent for $|t| < \varepsilon$, and the sum of the series is an analytic function of t . Moreover, the derivatives of the partial sums also converge uniformly to the derivative of the sum of the series.

From all this, it follows that the right-hand side of (3.14) is equal to

$$\left(\sum_{r=1}^{\infty} \frac{\partial}{\partial t} \sum_{\ell \in \mathcal{L}_r(G_\gamma)} w(\ell; x'_\gamma(t)) \Big|_{t=0} \right) \exp \left(\sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}_r(G_\gamma)} w(\ell; x'_\gamma(0)) \right).$$

In the first factor, the only loops that survive the differentiation are those that visit the edge uu^* , since only they contribute a factor t to the weight. Taking the derivative at $t = 0$, we are only left with those loops that visit the edge uu^* exactly once. In the second factor, because we set $t = 0$, the only loops that contribute are those that do not visit uu^* . This leaves precisely all loops in the graph G . The right-hand side of (3.14) therefore becomes

$$\left(\sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(G_\gamma): \\ \ell \text{ visits } uu^* \text{ once}}} w(\ell; x'_\gamma(1)) \right) \exp \left(\sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}_r(G)} w(\ell; x') \right),$$

from which, applying Theorem 1.2 again to the second factor, we find that

$$(3.15) \quad \langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \left(\sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(G_\gamma): \\ \ell \text{ visits } uu^* \text{ once}}} w(\ell; x'_\gamma(1)) \right) \frac{Z_G(x')}{Z_G(x)}.$$

Note the ratio of graph generating functions in (3.15). Recall that we have seen such a ratio of graph generating functions before in the low-temperature case, namely in (3.8). Thus, this ratio can be interpreted as a two-point function between the spins at u^* and v^* in a dual Ising model with positive boundary conditions at the dual low temperature β^* , given by $\exp(-2\beta^*) = \tanh \beta$. Using (3.9), we can express this ratio in terms of a sum over all u^*v^* -odd loops in the graph G , if we like. Hence, it is possible to express $\langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}}$ completely in terms of loops.

Next, we want to consider the limit as $G \rightarrow \mathbb{Z}^2$. By the argument given in Section 3.3, we already know that the ratio of graph generating functions in (3.15) converges to $\langle \sigma_{u^*} \sigma_{v^*} \rangle_{\mathbb{Z}^{2*}, \beta^*}^+$. It remains to consider what happens to the sum over the loops that visit uu^* once. To this end, for a general finite or infinite rectangle R in \mathbb{Z}^2 containing u and v , we define

$$(3.16) \quad a_r(R_\gamma; x'_\gamma) := \sum_{\substack{\ell \in \mathcal{L}_r(R_\gamma): \\ \ell \text{ visits } uu^* \text{ once}}} w(\ell; x'_\gamma),$$

where we have simplified the notation somewhat by writing $x'_\gamma := x'_\gamma(1)$.

As in the low-temperature case, the loops that contribute to $a_r(R_\gamma; x'_\gamma)$ must be confined to the box B_r^{uv} , defined in the same way as before, except possibly for the steps taken along the additional edges in E_γ , which do not count for the length of the loop, and are allowed to go outside B_r^{uv} . Hence,

$$(3.17) \quad a_r(R_\gamma; x'_\gamma) = a_r((R \cap B_r^{uv})_\gamma; x'_\gamma),$$

where $R \cap B_r^{uv}$ is the largest subgraph of R contained in B_r^{uv} , as before. It follows that $a_r(G_\gamma; x'_\gamma) \rightarrow a_r(\mathbb{Z}_\gamma^2; x'_\gamma)$ for all $r \geq 1$. As before, we now want to use dominated convergence to prove that

$$(3.18) \quad \lim_{G \rightarrow \mathbb{Z}^2} \sum_{r=1}^{\infty} a_r(G_\gamma; x'_\gamma) = \sum_{r=1}^{\infty} a_r(\mathbb{Z}_\gamma^2; x'_\gamma),$$

and that the right-hand side is absolutely summable. This requires an appropriate uniform bound (in R) on the right-hand side of (3.16).

To obtain this bound, by (3.17) it is sufficient to consider an arbitrary finite rectangle R in \mathbb{Z}^2 containing u and v , and such that R is contained in B_r^{uv} . Note that every loop of length r in R_γ which visits the edge uu^* once, has a representation of the form $\wp \oplus \gamma$, where \wp is a path of length r in R from v to u . Here, we use the facts that the part \wp of the loop never visits uu^* , and that the steps taken along γ from u to v do not contribute to the length of the loop, since they are along additional edges.

Let $\Lambda_R(x')$ be the transition matrix for the graph R with edge weights x'_e . Note that the sum of the weights of all paths \wp of length r from $v \rightarrow$ to $\uparrow u$, for instance, is given by the entry of the matrix $\Lambda_R^r(x')$ in row $v \rightarrow$ and column $\uparrow u$. To compute the sum of the weights of the corresponding loops $\wp \oplus \gamma$, we only need to multiply this entry by the factor $e^{i\phi/2}$, where ϕ is the sum of the turning angles encountered in the path from $\uparrow u$ to $v \rightarrow$ along the edges in E_γ . From these observations, we can conclude that

$$a_r(R_\gamma; x'_\gamma) \leq 16 \|\Lambda_R^r(x')\|_{\max},$$

where $\|\cdot\|_{\max}$ denotes the maximum-entry norm, and the factor 16 comes from the fact that there are 4 directed (representative) edges pointing out from v , and 4 pointing to u . By the proof of Theorem 1.3 and the fact that the maximum-entry norm of a matrix is bounded by the operator norm, and using that $(\tanh \beta_c)^{-1} = \sqrt{2} + 1$ by (1.14), we obtain

$$(3.19) \quad a_r(R_\gamma; x'_\gamma) \leq 16 \|\Lambda_R^r(x')\| \leq 16 \|\Lambda_R(x')\|^r = 16 \left(\frac{\tanh \beta}{\tanh \beta_c} \right)^r.$$

We emphasize that this bound holds uniformly for all finite and infinite rectangles R containing u and v . Hence, (3.18) holds by dominated convergence, and we have established the following intermediate result:

Theorem 3.4. *For all $\beta \in (0, \beta_c)$ and fixed $u, v \in \mathbb{Z}^2$ ($u \neq v$),*

$$\lim_{G \rightarrow \mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle_{G, \beta}^{\text{free}} = \left(\sum_{r=1}^{\infty} \sum_{\substack{\ell \in \mathcal{L}_r(\mathbb{Z}_\gamma^2): \\ \ell \text{ visits } uu^* \text{ once}}} w(\ell; x'_\gamma) \right) \langle \sigma_{u^*} \sigma_{v^*} \rangle_{\mathbb{Z}^{2*}, \beta^*}^+ =: \langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^{\text{free}}.$$

As an aside, we note that this result simplifies when u and v are on the same face of \mathbb{Z}^2 (i.e. $\|u - v\| = 1$), since then we can take $u^* = v^*$, so that $\sigma_{u^*} \sigma_{v^*} = 1$. Moreover, since the path γ is void in this case, none of the edge weights on G need to be modified, that is, $x' = x$ in this special case.

To conclude the proof of Theorem 1.6, we consider what happens to the two-point function when we let $\|u - v\|$ tend to infinity. Since the loops that visit uu^* necessarily have length at least $\|u - v\|$, we can write

$$\langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^{\text{free}} = \left(\sum_{r \geq \|u - v\|} a_r(\mathbb{Z}_\gamma^2; x'_\gamma) \right) \langle \sigma_{u^*} \sigma_{v^*} \rangle_{\mathbb{Z}^{2*}, \beta^*}^+.$$

By Theorem 1.5, the two-point function on the right is bounded between 0 and 1. Alternatively, at this stage we could also observe that the ratio of graph generating functions in (3.15) is always between -1 and $+1$, since the same even subgraphs contribute to both generating functions, but only in the numerator, some of them have a negative sign. Furthermore, the bound in (3.19) holds for $a_r(\mathbb{Z}_\gamma^2; x'_\gamma)$. Hence we have the explicit bound

$$|\langle \sigma_u \sigma_v \rangle_{\mathbb{Z}^2, \beta}^{\text{free}}| \leq 16 \sum_{r \geq \|u - v\|} \left(\frac{\tanh \beta}{\tanh \beta_c} \right)^r,$$

which tends to 0 exponentially fast as $\|u - v\| \rightarrow \infty$. \square

REFERENCES

- [1] M. Aizenman, D. J. Barsky, and R. Fernández, *The phase transition in a general class of Ising-type models is sharp*, J. Statist. Phys. **47** (1987), no. 3-4, 343–374. MR894398 (89f:82013)
- [2] V. Beffara and H. Duminil-Copin, *Smirnov's fermionic observable away from criticality*, 2010. arXiv:1010.0526 [math.PR].
- [3] P. N. Burgoyne, *Remarks on the combinatorial approach to the Ising problem*, J. Mathematical Phys. **4** (1963), 1320–1326. MR0155631 (27 #5565)
- [4] D. Cimasoni, *A generalized Kac-Ward formula*, J. Stat. Mech. Theory E. (2010JUL), P07023. arXiv:1004.3158.
- [5] G. A. T. F. da Costa and A. L. Maciel, *Combinatorial formulation of Ising model revisited*, Rev. Bras. Ens. Fis. **25** (2003), no. 1, 49–59. arXiv:math-ph/0304033.
- [6] N. P. Dolbilin, M. A. Shtan'ko, and M. I. Shtogrin, *Combinatorial problems in the two-dimensional Ising model*, Trudy Mat. Inst. Steklov. **196** (1991), 51–65. Translated in Proc. Steklov Inst. Math. 1992, no. 4, 57–72, Discrete geometry and topology (Russian). MR1111292 (92i:05083)
- [7] N. P. Dolbilin, Yu. M. Zinov'ev, A. S. Mishchenko, M. A. Shtan'ko, and M. I. Shtogrin, *The two-dimensional Ising model and the Kac-Ward determinant*, Izv. Ross. Akad. Nauk Ser. Mat. **63** (1999), no. 4, 79–100. MR1717680 (2000j:82008)
- [8] M. E. Fisher, *Critical temperatures of anisotropic Ising lattices. II. General upper bounds*, Phys. Rev. **162** (1967), no. 2, 480–485.

- [9] M. L. Glasser, *Exact partition function for the two-dimensional Ising model*, American Journal of Physics **38** (1970), no. 8, 1033–1036.
- [10] R. B. Griffiths, *Peierls proof of spontaneous magnetization in a two-dimensional Ising ferromagnet*, Phys. Rev. (2) **136** (1964), A437–A439. MR0189681 (32 #7103)
- [11] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Z. Physik **31** (1925FEB), 253–258.
- [12] M. Kac and J. C. Ward, *A combinatorial solution of the two-dimensional Ising model*, Phys. Rev. **88** (1952), no. 6, 1332–1337.
- [13] L. P. Kadanoff and H. Ceva, *Determination of an operator algebra for the two-dimensional Ising model*, Phys. Rev. B (3) **3** (1971), 3918–3939. MR0389111 (52 #9942)
- [14] B. Kaufman, *Crystal statistics. II. Partition function evaluated by spinor analysis*, Phys. Rev. **76** (1949), no. 8, 1232–1243.
- [15] B. Kaufman and L. Onsager, *Crystal statistics. III. Short-range order in a binary Ising lattice*, Phys. Rev. **76** (1949), no. 8, 1244–1252.
- [16] L. D. Landau and E. M. Lifshitz, *Course of theoretical physics. Vol. 5: Statistical physics, Part 1*, Translated from the Russian by J. B. Sykes and M. J. Kearsley. Third edition, revised and enlarged, Butterworth–Heinemann, Oxford, 1980.
- [17] B. M. McCoy and T. T. Wu, *The two-dimensional ising model*, Harvard University Press, Cambridge, Massachusetts, 1973.
- [18] L. Onsager, *Crystal statistics. I. A two-dimensional model with an order-disorder transition*, Phys. Rev. (2) **65** (1944), 117–149. MR0010315 (5,280d)
- [19] J. Palmer, *Planar Ising correlations*, Progress in Mathematical Physics, vol. 49, Birkhäuser Boston Inc., Boston, MA, 2007. MR2332010 (2008h:82001)
- [20] R. Peierls, *On Ising’s model of ferromagnetism*, Proc. Cambridge Phil. Soc. **32** (1936), 477–481.
- [21] S. Sherman, *Combinatorial aspects of the Ising model for ferromagnetism. I. A conjecture of Feynman on paths and graphs*, J. Mathematical Phys. **1** (1960), 202–217. MR0119512 (22 #10273)
- [22] ———, *Addendum: “Combinatorial aspects of the Ising model for ferromagnetism. I. A conjecture of Feynman on paths and graphs”*, J. Mathematical Phys. **4** (1963), 1213–1214. MR0156639 (27 #6560)
- [23] B. Simon, *The statistical mechanics of lattice gases. Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993. MR1239893 (95a:82001)
- [24] S. Smirnov, *Conformal invariance in random cluster models. I. Holomorphic fermions in the Ising model*, Ann. of Math. (2) **172** (2010), no. 2, 1435–1467. arXiv:0708.0039. MR2680496 (2011m:60302)
- [25] N. V. Vdovichenko, *A calculation of the partition function for a plane dipole lattice*, Soviet Physics JETP **20** (1965), 477–488. MR0181393 (31 #5622)
- [26] H. Whitney, *On regular closed curves in the plane*, Compositio Math. **4** (1937), 276–284. MR1556973

VU UNIVERSITY, DEPARTMENT OF MATHEMATICS, DE BOELELAAN 1081A, 1081 HV AMSTERDAM, THE NETHERLANDS

E-mail address: {w.kager,m.lis,r.w.j.meester}@vu.nl